

**Molecular Characterization Of The Activity And Requirements Of A
Novel And Promiscuous Bacteriophage Integrase**

Thesis submitted in accordance with the requirements of the University of
Liverpool for the Degree in Philosophy by

Mohammed R. Mohaisen

October 2017

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

| | |
|--|-----------|
| <i>1.1 Viruses.....</i> | <i>1</i> |
| 1.1.1 Bacteriophages..... | 3 |
| 1.1.2 Temperate Bacteriophages..... | 6 |
| <i>1.2 Lambda phage, the archetypal lambdoid phage.....</i> | <i>7</i> |
| 1.2.1 Lambda lytic cycle..... | 10 |
| 1.2.2 Lambda lysogenic pathway..... | 11 |
| 1.2.3 Lambda phage induction..... | 15 |
| <i>1.3 Temperate Phage and HGT.....</i> | <i>19</i> |
| 1.3.1 Phage Transduction..... | 20 |
| 1.3.2 Phage Conversion..... | 24 |
| <i>1.4 Site-specific recombination.....</i> | <i>29</i> |
| 1.4.1. Serine integrases..... | 33 |
| 1.4.2 Applications of serine site-specific recombinases..... | 38 |
| 1.4.3 Tyrosine recombinases..... | 40 |
| 1.4.4 Applications of tyrosine recombinases..... | 46 |
| 1.4.4.1 Gateway® cloning system..... | 46 |
| 1.4.4.2 Chromosomal integration..... | 49 |
| 1.4.4.3 Tyrosine integrases in mammalian cells..... | 51 |
| <i>1.5. Φ24B Integrase Protein.....</i> | <i>53</i> |
| <i>1.6. The study aims.....</i> | <i>56</i> |

CHAPTER 2: GENERAL MATERIAL AND METHODS 58

| | |
|---|-----------|
| <i>2.1. Bacterial Strains.....</i> | <i>58</i> |
| <i>2.2. Plasmids, culture conditions, and buffers used in this study.....</i> | <i>62</i> |
| <i>2.3. Bacterial genomic DNA Extraction.....</i> | <i>62</i> |

| | |
|--|----|
| 2.4. Plasmid recovery..... | 63 |
| 2.4.1 Plasmid mini prep..... | 63 |
| 2.4.2. Plasmid midi prep..... | 63 |
| 2.5. Polymerase chain reaction (PCR)..... | 64 |
| 2.6. Agarose gel electrophoreses..... | 72 |
| 2.7. Recovery of the DNA from the agarose gel..... | 73 |
| 2.8. Restriction endonuclease digestion..... | 73 |
| 2.9. Ligation of DNA fragments..... | 73 |
| 2.10. Production of electrocompetent cells..... | 73 |
| 2.11. Transformation of electrocompetent cells..... | 74 |
| 2.12. Preparation of chemically competent cells..... | 74 |
| 2.13. Transformation of chemically competent cells..... | 75 |
| 2.14. Colony PCR..... | 75 |
| 2.15. Protein quantification: Bicinchoninic Acide assay (BCA)..... | 76 |
| 2.16. The Sequencing analysis..... | 76 |

CHAPTER 3: IN VITRO RECOMBINATION 80

| | |
|---|----|
| 3.1. Background..... | 80 |
| 3.1.1. Site-Specific Recombinases..... | 80 |
| 3.1.2. Lambda phage Integrase..... | 80 |
| 3.1.3. Integrase of $\Phi 24_B$ | 86 |
| 3.1.4 Aims..... | 89 |
| 3.2. Specific Methods..... | 90 |
| 3.2.1. Cloning of bacterial and phage attachment sites..... | 90 |
| 3.2.2. Protein Overexpression..... | 90 |
| 3.2.3. SDS Gel PAGE..... | 93 |
| 3.2.4. Simi-Dry Western blot..... | 94 |

| | |
|---|------------|
| 3.2.5. Protein purification..... | 95 |
| 3.2.6. <i>In Vitro</i> Recombination Assay..... | 97 |
| 3.3. <i>Results</i> | 99 |
| 3.3.1 IHF production..... | 99 |
| 3.3.2 Int _{Φ24B} Production..... | 101 |
| 3.3.3 Production of attP Φ24B, attB Φ24B constructs..... | 102 |
| 3.3.5 Integration Assay (<i>in vitro</i>)..... | 104 |
| 3.4. <i>Discussion</i> | 108 |
| CHAPTER 4: <i>IN SITU</i> RECOMBINATION ASSAY | 111 |
| 4.1. <i>Background</i> | 111 |
| 4.1.1. The Stx Phage Φ24 _B | 115 |
| 4.1.2. <i>In situ</i> assay..... | 116 |
| 4.1.3 Aims..... | 119 |
| 4.2. <i>Specific Methods</i> | 120 |
| 4.2.1. Plasmids Cloning of <i>attB</i> and <i>attP</i> sites..... | 120 |
| 4.2.2. <i>In situ</i> recombination assay..... | 124 |
| 4.2.3. The importance of IHF for Φ24B integration..... | 125 |
| 4.3. <i>Results</i> | 126 |
| 4.3.2. The minimal necessary <i>attP</i> size..... | 126 |
| 4.3.2. The minimal necessary <i>attB</i> size..... | 127 |
| 4.3.3 IHF importance determination..... | 132 |
| 4.4. <i>Discussion</i> | 132 |
| CHAPTER 5: IDENTIFICATION OF INTEGRATION SITE PREFERENCES | 137 |
| 5.1 <i>Background</i> | 137 |
| 5.2. <i>Specific Methods</i> | 143 |
| 5.2.1. Bacterial strains, plasmids and media..... | 143 |

| | |
|--|------------|
| 5.2.2. PCR cloning..... | 143 |
| 5.2.3. <i>In situ</i> recombination assay..... | 144 |
| 5.2.4. Comparative qPCR assay..... | 147 |
| 5.3. Results..... | 149 |
| 5.3.1. Relative quantitative recombination events..... | 149 |
| 5.3.2. Recombination site preference of <i>attP</i> site..... | 152 |
| 5.4. Discussion..... | 153 |
| Chapter 6: GENERAL DISCUSSION AND FUTURE WORK | 158 |
| 6.1. General Discussion..... | 158 |
| 6.2 Future Work..... | 166 |
| 6.2.1 Synthetic Recombination Vectors..... | 166 |
| 6.2.1.1 Integration and DNA homology..... | 166 |
| 6.2.1.2 Gibson Assembly..... | 170 |
| 6.2.1.3 BioBricks..... | 173 |
| 6.2.1.4 BioBrick Standard Assembly..... | 175 |
| 6.2.2. Proof of concept: synthetic recombination vector assay..... | 178 |
| 6.2.3 Necessary Integration Host Factors..... | 186 |
| 6.2.3.1 Protein fractionation..... | 186 |
| 6.2.3.2 Proof of concepts..... | 187 |
| CHAPTER 7: REFERENCES | 188 |

TABLE OF FIGURES

| | |
|--|----|
| Figure 1.1. Schematic diagram explains transcription of virus genomes | 2 |
| Figure 1.2. The temperate bacteriophage lytic and lysogenic cycle | 4 |
| Figure 1.3. Lambda genome and transcription map | 9 |
| Figure 1.4. The lambda immunity region | 14 |
| Figure 1.5. Structure and localization of λ phage | 17 |
| Figure 1.6. The membrane fusion steps model for Rz-Rz1 lytic function | 18 |
| Figure 1.7. The specialized transduction by lambda phage | 23 |
| Figure 1.8. Lambda phage site-specific integration and excision | 30 |
| Figure 1.9. Overview of the cleavage and rejoining of attachment sites by serine recombinase. | 37 |
| Figure 1.10. DNA recombination by lambda integrase and most tyrosine recombinases | 42 |
| Figure 1.11. Protein domain structure of λ Int. | 45 |
| Figure 1.12. Gateway® cloning system | 47 |
| Figure 1.13. The integrase region in the ϕ 24B phage genome | 55 |
| Figure 2.1. Site-directed mutagenesis technique | 78 |
| Figure 3.1. Circular lambda phage DNA is converted to an integrated prophage by a reciprocal recombination between <i>attP</i> and <i>attB</i> . | 81 |
| Figure 3.2: Lambda <i>attP</i> and <i>attB</i> sites | 85 |
| Figure 3.3: in vitro recombination assay using <i>pattB</i> and <i>pattP</i> plasmids and required proteins in a final condition of 50 μ l of suitable buffer. | 98 |

| | |
|--|------------|
| Figure 3.4: The comparison between IHF overexpressions | 100 |
| Figure 3.5: Production of Integrase from a His-tagged cloned. | 101 |
| Figure 3.6. Colony PCR of the <i>E. coli</i> Top 10 transformants to confirm the construction of pPCR2.1-attP200 bp. | 103 |
| Figure 3.7. Colony PCR of the <i>E. coli</i> Top 10 transformants to confirm the construction of pPCR2.1-attP300 bp. | 104 |
| Figure 3.8: Quality control for purified IHF and Integrase. | 105 |
| Figure 3.9: <i>In vitro</i> recombination tests using cells crude extract. | 107 |
| Figure 3.10: PCR assay of the recombination products. | 108 |
| Figure 4.1. The Int/DNA intasome complex based on the models | 114 |
| Figure 4.2. The identification of bacterial and bacteriophage integration attachment sites. | 116 |
| Figure 4.3. Restriction free cloning | 123 |
| Figure 4.4. <i>In situ</i> recombination assay. | 125 |
| Figure 4.5. colony PCR of plasmids in the <i>E. coli</i> Top 10 transformants confirming the transformation of <i>pattP</i> (600 bp) to cells already harbouring pΦ24_B-int and <i>pattB</i>. | 126 |
| Figure 4.6. PCR detection of the resultant attRL hybrid site formation from minimal attP sequence detection assays. | 127 |
| Figure 4.7. PCR results from the <i>E. coli</i> Top 10 transformants to confirm the presence of pACYC-Duet-<i>attB</i> 600 bp. | 128 |
| Figure 4.8. PCR detection of the resultant attRL sites from minimal attB sequence <i>in situ</i> integration assays. | 130 |
| Figure 4.9. Comparison between the essential sequences of the λ <i>attB</i> site & Φ24_B<i>attB</i> sites. | 131 |

| | |
|---|-----|
| Figure 4.10. <i>In situ</i> recombination assay using the <i>E coli</i> strain JW1702-1. | 132 |
| Figure 5.1: The alignment of four distinct bacterial integration sites in same genome. | 140 |
| Figure 5.2. The substrates and the final recombination products of <i>in situ</i> recombination assay. | 146 |
| Figure 5.3. The relative quantitative recombination events in the secondary sites mediated by $\Phi 24_B$ Int measured by qPCR as compared to the use of attB _{intS} . | 151 |
| Figure 5.4. The attB secondary sites recombination preference, measured by qPCR time course. | 153 |
| Figure 6.1. The effect of DNA sequence homology in the lambda attB and attP site recombination outcome mediated by lambda Int. | 168 |
| Figure 6.2. Gibson assembly overview. | 171 |
| Figure 6.3. A schematic diagram depicting the assembling of two BioBricks using BioBrick Standard Assembly. | 177 |
| Figure 6.4. The synthesized integration vector. | 179 |
| Figure 6.5. Bacterial growth of Top10 <i>E coli</i> cells following transformation of the integration vector | 180 |
| Figure 6.6. Recombination product following plasmid harvest, restriction enzyme digestion and electrophoresis gel run. | 182 |
| Figure 6.7. Bacterial colonies that resulted after the transformation of the purified recombined vector to New Top10 competent cells. | 183 |
| Figure 6.8. A schematic diagram explains the recombination events between attP and attB sites within the synthesized integration vector. | 184 |
| Figure 6.9. Positive control vector demonstrating the ability to capture integration events. | 185 |

TABLE OF TABLES

| | |
|--|------------|
| Table 1.1: Selected, well characterised site-specific recombinases and their requirements for recombination | 35 |
| Table 2.1.1. Bacterial strains used in this study | 58 |
| Table 2.2.1. Plasmids used in our study | 59 |
| Table 2.2.2. Composition of growth media, buffers and antibiotic preparations | 61 |
| Table 2.5.1. PCR mastermix using Q5 high fidelity DNA polymerase | 66 |
| Table 2.5.2. Oligonucleotide primers used for PCR amplifications of DNA | 66 |
| Table 3.2.1 Western Blot Buffers | 94 |
| Table 6.1. The recombination vectors that we created for synthetic recombination derived by Φ24B Int. | 180 |

ABBREVIATION

| | |
|-------------------------|---|
| Φ | bacteriophage |
| λ | lambda |
| attB | Bacterial attachment site |
| attL | Hybrid attachment site-Left |
| attP | Phage attachment site |
| attR | Hybrid attachment site-Right |
| Amp | Ampicillin |
| Cat | Chloramphenicol acetyl transferase |
| C_T | Cycle threshold |
| CTX | Cholera toxin phages |
| ddH₂O | Double distilled water |
| dsDNA | Double stranded DNA |
| DNA | Deoxyribonucleic acid |
| DT | Diphtheria toxin |
| DTT | Dithiothreitol |
| EDTA | Ethylene diamine tetra acetic acid |
| EHEC | Enterohaemorrhagic <i>Escherichia coli</i> |
| Fis | Factor for Inversion Stimulation |
| HUS | Haemolytic Uraemic Syndrome |
| IM | Inner membrane |
| Int | Integrase |
| int | Integrase encoding gene |
| IHF | Integration host factor |
| IPTG | Isopropyl-D-1-thiogalactopyranoside |

| | |
|----------------|--|
| Kan | Kanamycin |
| LB | Luria Bertani |
| LPS | Lipopolysaccharide |
| mRNA | Messenger RNA |
| NCBI | National Centre for for Biotechnology Information |
| OD | Optical density |
| OM | Outer membrane |
| ORF | Open Reading Frame |
| P | Promoter |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain Reaction |
| Phage | Bacteriophage |
| Phusion | Recombinant hi-fidelity DNA polymerase |
| qPCR | Quantitative (real-time) PCR |
| RNA | Ribonucleic acid |
| rRNA | Ribosomal ribonucleic acid |
| SDS | Sodium Dodecyl Sulphate |
| Spec | Spectinomycin |
| Spp. | Species |
| STEC | Shigatoxigenic <i>E coli</i> |
| Stx | Shiga toxin |
| ssDNA | Single-stranded DNA |
| t | Transcription terminator site |
| T7 | Bacteriophage T7 |
| TAE | Tris-acetate EDTA |

| | |
|--------------|---|
| Taq | Recombination DNA polymerase from <i>Thermus aquaticus</i> |
| TE | Tris EDTA |
| TEMED | Tetramethylethylenediamine |
| Tris | Tris(hydroxymethyl)methylanime |
| TTP | Thrombotic Thrombocytopenic purpura |
| U | Units |
| UV | Ultra-violet |
| Xis | excisionase |

UNITES

| | |
|--------------|------------------------------|
| bp | Base pair |
| c.f.u | Colony forming units |
| cm | Centimetres |
| g | Grames |
| h | Hours |
| kb | Kilobase |
| kDa | Kilodaltons |
| L | Litre |
| M | Moles per litre |
| MA | Milliamps |
| mg | Milligrams |
| ml | Millilitres |
| mM | Millimoles per litre |
| mm | Millimetre |
| mV | Millivolts |
| MW | Molecular Weight |
| ng | Nanograms |
| °C | Degrees Centigrade |
| OD | Optical density |
| r.p.m | Revolution per minute |
| µg | Micrograms |
| µM | Micromoles per litre |
| V | Volts |
| v/v | Volume/volume |
| w/v | Weight/volume |

ACKNOWLEDGEMENTS

Foremost, I would like to thank my God (Allah) for making this work possible and giving me the ability and patience to complete my research at the wonderful Department of Functional and comparative genomics, Institute of Integrative biology, University of Liverpool. Thank God for allowing me study under great supervision and surrounding me with helpful people throughout my study.

I would like to thank my supervisors Dr Heather Allison and Professor Alan McCarthy for all their invaluable support, expert advice, patience and assistance throughout my PhD. I have no words to explain my gratefulness to them. I would also like to thank the Higher Committee for Education Development in Iraq for their generous funding of my PhD research.

I would like to thank my parents who proud of me and who never forget to pray for me to finish my study. I cannot forget to thank my lovely family, my wife and the three kids, who share this life with me and make it full of success and happiness.

I would like to thank all my friends in lab H, they have all been so kind and helpful. Thanks a lot for welcoming me from the first day and making these four years were the best years of life. I will not forget these wonderful people.

Abstract

Stx bacteriophages are responsible for the dissemination and production of Shiga toxin genes (stx) across the Shigatoxigenic *E. coli* (STEC). These toxigenic bacteriophage hosts can cause severe, life-threatening illness, and Shiga toxin (Stx) is responsible for the severe nature of EHEC infection, a subset of pathogenic STEC. At the point of Stx phage infection, the injected phage DNA can direct its integration into the bacterial chromosome becoming a prophage; the host cell is then known as a lysogen. Unusually, our model Stx phage, Φ 24B, can integrate into at least four distinct sites within the *E. coli* genome that shared no easily identifiable recognition sequence pattern. The identification of what are actually required for phage and bacterial DNAs recombination has been tested using both *in vitro* and *in situ* recombination assays. These assays enabled the simple manipulation of bacterial attachment site (*attB*) and phage attachment site (*attP*) sequences.

The aim of the study is to fully characterize the requirements of this promiscuous integrase, carried by the Stx phage Φ 24B (Int Φ 24B), to drive integration. These assays enabled us to identify the minimal necessary flanking sequences for *attB* site identified (21 bp and 49 bp from the right and left the cross over region, respectively) and the *attP* site (200 bp each side). Furthermore, we identified that the Φ 24B integrase does not need Integration Host Factor (IHF) to drive integration.

Finally, as this integrase can integrate the phage genome inside at least four different bacterial attachment sites (*attBs*), these sites were identified, sequenced and cloned in different compatible plasmids to be transformed to one cell, and the frequency and preference of each recombination were tested by means of qPCR. The results showed, the recombination with secondary attachment sites was less frequent than that with the primary site. Furthermore, within five minutes, there was also a preference of site, as recombination occurred in both *attB*₁ and *attB*₂, while recombination with the *attB*₄ and *attB*₃ sites, took 1 hour and 2 hours, respectively.

CHAPTER 1. INTRODUCTION

1.1 Viruses

Viruses are the most abundant biological agents on the planet (Wilhelm and Suttle 1999). They cannot replicate and produce new progeny without first infecting and replicating inside another organism's living cells. They can infect all types of life forms from plants and animals to microorganisms like protozoa, bacteria and archaea (Lawrence et al., 2009). Every studied cellular organism has its own viruses or virus-like genetic elements (Fauquet and Fargette, 2005). Since they were first described in 1892 (Iwanowski, Sept. 1892), more than 5000 viral species have been described in details (Breitbart and Rohwer, 2005). Viruses possess a variety of genetic replication and expression mechanisms (Fig.1.1) in contrast to the uniformity of the cellular genetic cycles (DNA \Rightarrow RNA \Rightarrow Protein) (Carter and Saunders, 2007).

When the virus is not inside an infected living cell, they exist in the form of particles known as virions, which are composed of two or three parts; double stranded or single stranded, DNA or RNA genomic material and the capsid, which is a protein coat that surrounds and protects the genetic material. In some cases, viruses have an envelope of lipids that surrounds the inside or outside of the capsid when the virus is outside their host cells

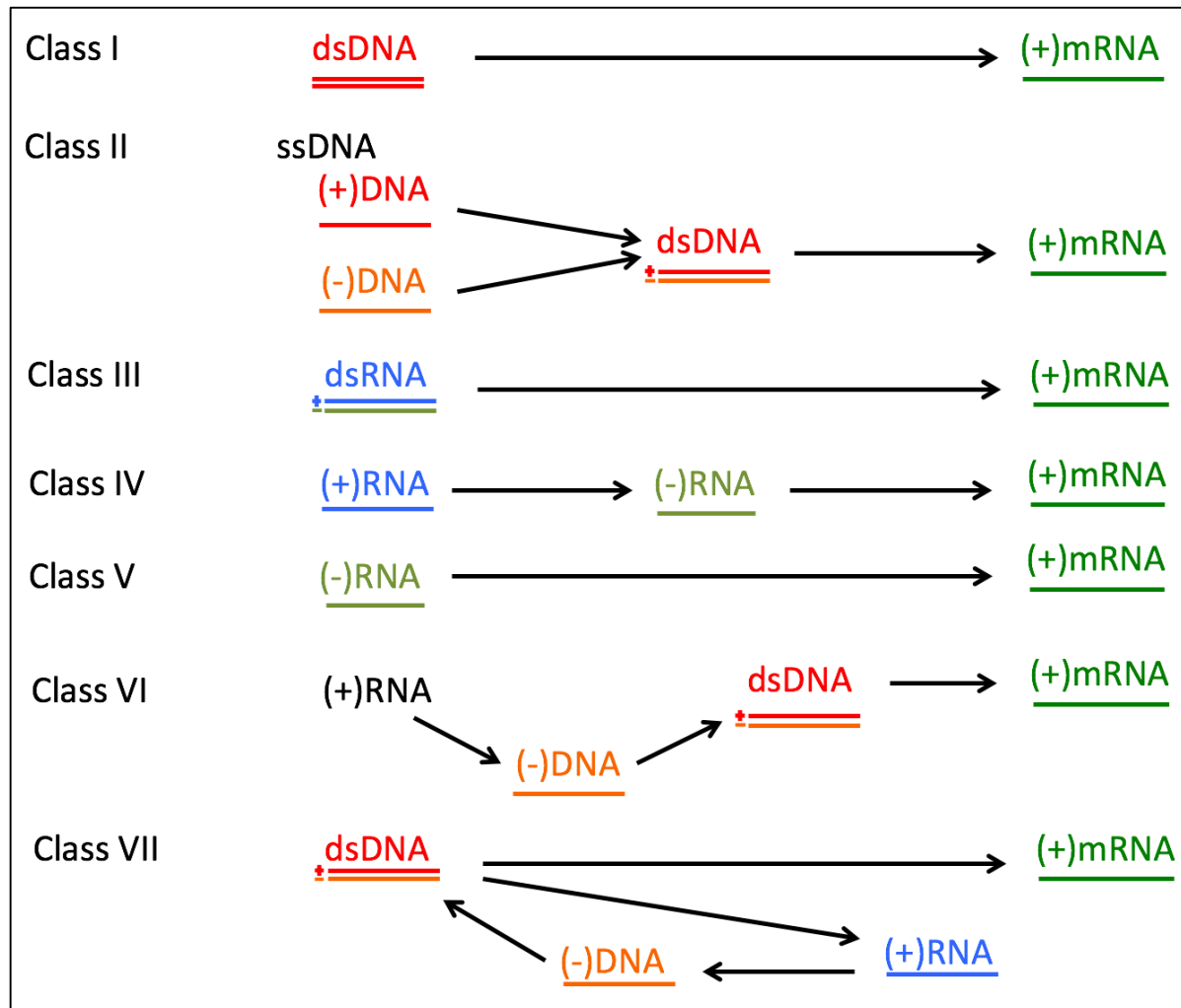


Figure 1.1. Schematic diagram explaining transcription of virus genomes. According to the difference in their genome composition, transcription of the viral genome can be classified into seven classes. The class I viruses have dsDNA genome (Red color) and transcription is similar to that of the higher organisms that have dsDNA. The genomes of most of these viruses have open reading frames (ORFs) in both directions, which is the reason behind why (-) and (+) are not indicated for dsDNA. Class II have genomes with either (+) which means they have same sequence as the mRNA (except that thymine replaces uracil), or (-) (orange color) which means these viruses have ssDNA sequence complementary to the mRNA. Class IV Viral genomes with (+) RNA have the same sequence as the mRNA, while classes V viruses with (-) RNA have the sequence complementary to the mRNA. While both of (-) and (+) RNAs sequences are present in class III dsRNA (blue color) genome viruses. Class VI viruses genome have the same sequence as the mRNA, however, unlike class IV (+)RNA, its genome transcribes to give (-)DNA in stead of (-)RNA in class IV. In Class VII genome transcription, the (+) RNA shown in blue color is a pregenome RNA, functions as a template for DNA synthesis.

(Adapted from Carter and Saunders, 2007)

(Prescott et al., 2002). Virus particles are diverse in their shapes, which range from simple helical and icosahedral forms for some viral species to more complicated structures for other viruses. Most virus species have virions that are too small to be seen with an optical microscope, their sizes ranging from 20-300 nm in diameter (Prescott et al., 2002).

In summary, viruses differ from other eukaryotic or prokaryotic living cells in at least three things: 1) their simple non-cellular organization, 2) the presence of either DNA or RNA as genetic material but not both and 3) they are unable to replicate independently of a host cell as eukaryotes and prokaryotes do (Prescott et al., 2002).

1.1.1 Bacteriophages

The majority of viruses are that infect and replicate inside bacterial cells (Hambly and Suttle, 2005). These viruses are called bacteriophages or phages for short. Phage infection, for all bacteriophages, can lead to the production of progeny phages via the lytic replication cycle. Lytic replication describes the process by which the phage uses the bacterial host cell as a factory in order to produce new phage capsids, tails and other structural proteins to form new phage particles within the bacterium (Fig. 1.2). These particles are often released by phage-directed lysis of the host cell, resulting in the escape of ~100 progeny phages per infected host cell (Bradley, 1967). However, not all bacteriophage immediately

replicate upon entering a host cell. Some phage, like P1 phage, inject their genome into their host cell where the genome becomes an extrachromosomal episome capable of self replication and thus part of the genome of the infected host cell (Ikeda and Tomizawa, 1968)

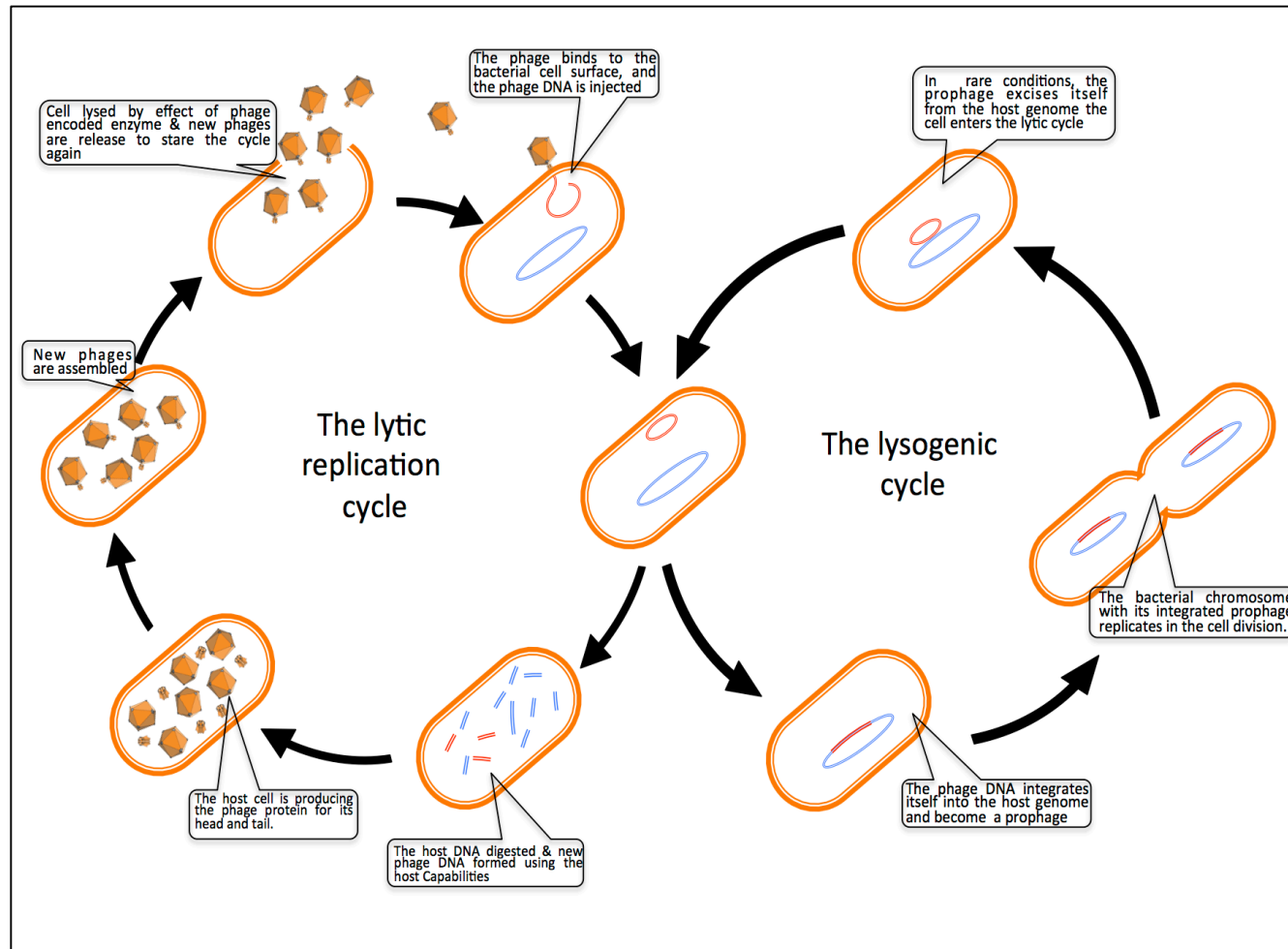


Figure 1.2. The temperate bacteriophage lytic and lysogenic cycle. There are two pathways that generally govern bacteriophage replication once the viral genome has entered the bacterial host cell. 1) The lytic cycle (left side) uses the bacterial cell as a factory in order to produce new phage capsids, tails, other structural proteins and new viral genome copies in order to form new phage particles within the bacterium. If the host cell is *E. coli* this process can take around 45 minutes before phage proteins direct the lysis of the host cell, releasing ~100 progeny phages. 2) The lysogenic replication cycle (right side) can also occur. The entering phage genome integrates itself within the bacterial genome, becoming a prophage and changing its host cell into a lysogen. The health of the host cell and the ratio of phages to host cell numbers at the point of infection can influence which pathway is taken, though the lytic cycle is more frequently taken by Lambda phage (Adapted from Campbell, 2003).

Sometimes the injected phage DNA directs its integration into the host cell's genome where it becomes known as a prophage, and its host cell becomes known as a lysogen. Phages capable of becoming a stable part of the host cell's genome are known as temperate phage (Hendrix, 1983). The prophage is passed along to other daughter cells of the lysogen and thus the prophage replicates passively, but concomitantly with the lysogen (Fig. 1.2).

1.1.2 Temperate Bacteriophages

Temperate phages are capable of entering either the lytic or lysogenic life cycle once their DNA is injected in the host cytoplasm (Fig. 1.2)(Kaiser, 1957). Temperate phage carry at least one gene capable of stabilizing the prophage sequences within the lysogen's genome (Hendrix, 1983). And then they rely on the activity of prophage encoded repressors to block the viral genes involved in lytic replication(Madsen et al., 1999). In the lysogen, with the exception of one prophage gene, all prophage gene expression is turned off. The protein product of this single expressed gene provides immunity to superinfection against incoming homoimmune phages by ensuring all genes in the incoming phage are repressed. The prophage genome is passively replicated as part of the lysogen's genome and distributed to the offspring bacterial cells as the lysogen grows and divides (Fig. 1.2). Though it is possible to turn on the lytic replication mechanism of the prophage, a process known as induction. The prophage is rarely induced in the lysogenic daughter cells, and the passive replication of the prophage might be

repeated indefinitely (Hendrix, 1983). The lytic and lysogenic pathways are controlled by a number of specific factors that must be synthesized in the phage-infected host to direct either lytic or lysogenic replication of the infecting phage.

1.2 Lambda phage, the archetypal lambdoid phage

Lambda phage, λ , the archetype of phages within the lambdoid family, was the first temperate phage discovered in the early 1960s (Hendrix, 1983). Its regulatory and replication mechanisms have been well understood following more than 60 years of extensive studies (Court et al., 2007). This bacteriophage being a temperate phage, follows either a lytic or lysogenic lifecycle inside its bacterial host cell, *Escherichia coli*. As a result of all of this work, λ has become useful and important to understanding molecular microbiology fields; like detection of frameshift mutation (Hendrix and Duda, 1992), DNA recombination (Meselson and Weigle, 1961; Signer and Weil, 1968), and Horizontal gene transfer (Young et al., 1979).

Lambda phage possesses an icosahedral head (capsid) of about 50-60 nm in diameter. The components of this capsid are encoded by ten head genes (*NuI*, *A*, *W*, *B*, *C*, *Nu3*, *D*, *E*, *FI*, and *FII*) (Fig.1.3). The phage also possesses a long, flexible tail of about 150 nm length, which is encoded by 12 genes (*Z*, *U*, *V*, *G*, *T*, *H*, *M*, *L*, *K*, *I*, *J*, *lo* and *stf*) (Fig.1.3). The Lambda genome is comprised of double-stranded DNA of ~48.5 kb. This DNA is linear when packaged into the phage's capsid and

circularises once it is injected into the host cell's cytoplasm by the action of a bacterial enzyme (*E. coli* DNA ligase in case of lambda phage using the cohesive ends found on the lambda genome)(Gellert, 1967).

When λ phage genomic DNA enters the host cell, both the *N* and *cro* genes are immediately transcribed, after the binding of bacterial RNA polymerase to the promoters P_R and P_L in the phage chromosome (Fig. 1.3). The protein N is called an anti-terminator as it effectively turns on genes in the phage DNA immediately downstream of *N* and *cro* by promoting the extension of the mRNAs of *N* and *cro* beyond the transcription termination signals that lie at the 3' ends of these 2 genes. The N protein is able to extend the length of the transcripts begun at P_R and P_L due to its ability to bind specific sequences called *Nut* (Nutilisation).

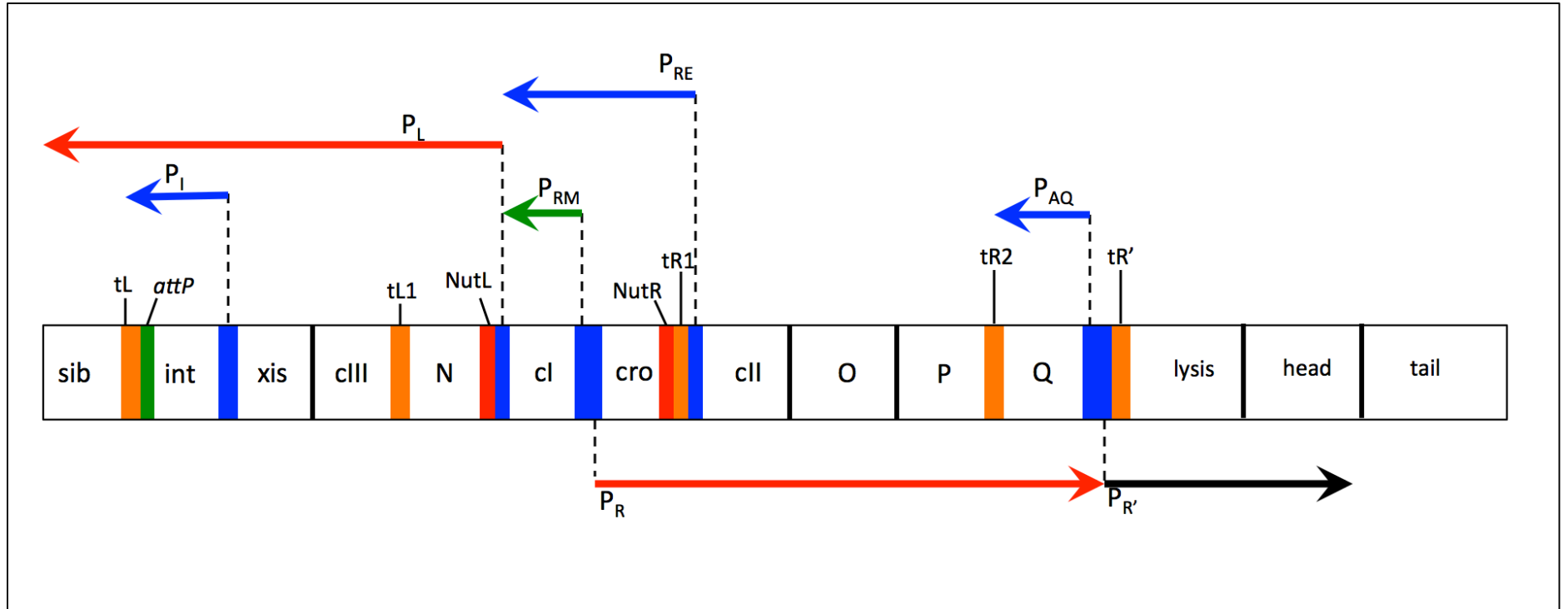


Figure 1.3. Lambda genome and transcription map. There are two Nut sites (red color), one is located between P_L promoter and *N* gene, and the second one is located just to the right of *cro* gene. After the *N* protein is produced, it binds to Nut sites in order for transcription to proceed. This transcription is enabled beyond the terminators *tR1* and *tL1*. The early transcripts from P_L and P_R are presented as red arrows.

Due to the actions of the *N* protein, *cIII*, *xis* and *int* genes are turned on from P_L , and *cII*, *O*, *P*, and *Q* genes are turned on from P_R .

The terminators *tL1*, *tL*, *tR1*, *tR2*, *tR'* are labeled with red color. The late transcription from $P_{R'}$ is shown as black arrow. The P_{RM} transcript activated by *CI* is presented by green arrow.

Transcription activated by *CII* from the promoters P_I , P_{RE} and P_{AQ} are shown with blue arrows (Adapted from Court et al., 2007).

When the N protein is bound to the *Nut* sites RNA polymerase can pass over the terminator sequences that would otherwise terminate transcription. The *Nut* sequences are located between P_L and the end of the *N* gene, and the second one is located just to the right of *cro* (Oppenheim et al., 2005; Friedman and Court, 1995). In the early infection stages, and due to the action of the N protein, the genes *cIII*, *xis* and *int* are turned on from P_L , and *cII*, *O*, *P*, and *Q* genes are turned on from P_R . The balance between expression and activities of these various gene products then determines whether the infecting phage will replicate according to the lytic or lysogenic pathways (Fig 1.3).

1.2.1 Lambda lytic cycle

The lytic replication cycle of lambda phage becomes the preferred pathway when high levels of Q protein are expressed from the P_R promoter, which enables the activation of the late gene promoter by Q. The late gene promoter, controlled by Q, is responsible for the production of the structural proteins, which comprise the capsid and tail structures and the expression of the proteins responsible for lysis of the host cell (*S*, *R*, *Rz* and *Rz*) (Fig.1.3). The production of Q has to reach a threshold level to initiate the lytic cycle. At that point, the Q protein, at sufficient levels, works as an anti-terminator for *E. coli* RNA polymerase at *qut* sites (Q utilization, similar to *Nut sites*), which lie immediately downstream of the *Q* gene, at an area called P_R' . The promoter P_R' enables transcription the phage late genes and enables subsequent host cell lysis (Roberts et al., 1998).

At the same time, the expression of the CI repressor is prevented due to the binding of Cro dimers to the operator binding sites O_{R3} and O_{R2} . Cro protein also binds to O_{L2} and O_{L3} to turn off the P_L promoter, shutting off all transcription in the leftward direction. At high concentrations, Cro protein can also bind to O_{R1} preventing further Cro expression from P_R . conditions that result in this very high concentration of Cro are the same conditions that enable the amount of Q to be sufficient to activate the production of late genes. The transcription of genes, as a result of Q activation from P_R' , is faster than the transcription of genes activated by N (Yang et al., 1987; Kobiler et al., 2005). Newly replicated phage genome copies are then encased into the newly synthesized capsids, tails are then added and the phage directs the lysis of the host cell with S, Rz and Rz1, releasing all of the newly synthesized phage progeny (Hendrix, 1983).

1.2.2 Lambda lysogenic pathway

The two-phage proteins necessary for the formation of stable lysogens are integrase (Int) and repressor (CI). The expression of these proteins is dependent on the activity of another protein, CII, which is expressed the early in the infection process from P_R after N alleviates the tR1 terminator (Fig. 1.3). CII is a transcription activator that encourages the host RNA polymerase to initiate transcription at three promoters, P_{Int} , P_{RE} and P_{AQ} (Oppenheim et al., 2005). P_{Int} , P_{RE} and P_{AQ} are phage promoters that control the transcription of integrase,

repressor and anti-sense Q transcripts, respectively. In order to activate the expression from these promoters, CII needs to accumulate to a threshold level within the infected cell's cytoplasm. The accumulation of CII is limited by a host cell encoded protease, HflB (also known as FtsH). HflB is capable of destroying CII by cleaving the carboxy terminus of CII. The phage attempts to limit this degradation by the expression and action of the CIII protein. CIII inhibits HflB activity (Shotland et al., 1997; Halder et al., 2008; Herman et al., 1997). The levels of HflB are controlled by the health of the bacterial cell. So only conditions that inhibit the expression of bacterial HflB (*e.g.* low nutrient availability, sub-optimal growth conditions, etc) result in conditions where CII survives and acts upon the three promoters, P_{Int}, P_{RE} and P_{AQ} (Kobiler et al., 2007; Kobiler et al., 2002).

When Int protein is produced it functions as an enzyme that is capable of catalyzing site-specific recombination between the phage and bacterial host's genomes (Int function will be discussed in 1.4.1 Tyrosine recombinases). The gene *int* is actually transcribed from two different promoters P_L and P_{Int} (Schindler and Echols, 1981). Transcription that begins at P_{Int} is actually activated by the CII protein and stops at the tL termination site (Fig. 1.3). While transcription that is initiated from the P_L promoter passes over the tL1 site (due to the function of N) and transcribes the *sib* site which is located downstream of *int* (Fig. 1.3). The *sib* site forms a large loop structure on the mRNA that serves as a target for bacterial RNase III. RNaseIII cleaves the loop leaving the end 3' of the *int* gene transcript

susceptible to degradation by polynucleotide phosphorylase (PNPase), that stops *int* transcription (Hendrix, 1983).

In order to stop expression of the late genes, CII activates the P_{AQ} promoter. The product of the P_{AQ} promoter checks the function of Q protein by creation of a Q antisense RNA. This antisense transcript binds to the complementary 5' end of the Q transcript, resulting in the destruction of the Q transcript.

The third promoter activated by CII is P_{RE} (Fien et al., 1984). This promoter, named the promoter for repressor establishment, is responsible for enabling the expression of CI, the lambda repressor. The CI protein shuts off all gene expression from both P_R and P_L by binding to their operator sites. The right operator region (O_R), harbors 2 divergent promoters (Fig. 1.3), P_{RM} and P_R whose activities are controlled by regulator proteins (Cro and CI) that bind specific combinations of binding sites known as O_{R1} , O_{R2} and O_{R3} . There is also a slightly more simple operator region, O_L which harbours one promoter, P_L whose transcriptional activity is controlled by three operator binding sites, O_{L1} , O_{L2} and O_{L3} . Two CI dimers bind cooperatively to O_{R1} and O_{R2} to repress P_R , and to repress P_L , another two CI dimers bind O_{L1} and O_{L2} (Dodd et al., 2004). The two tetramers of CI are then capable of binding other CI dimers, creating a loop between both operators (Fig. 1.4). At this point, due to the repression of P_R promoter, there is no more CII production and CI transcription from P_{RE} stops.

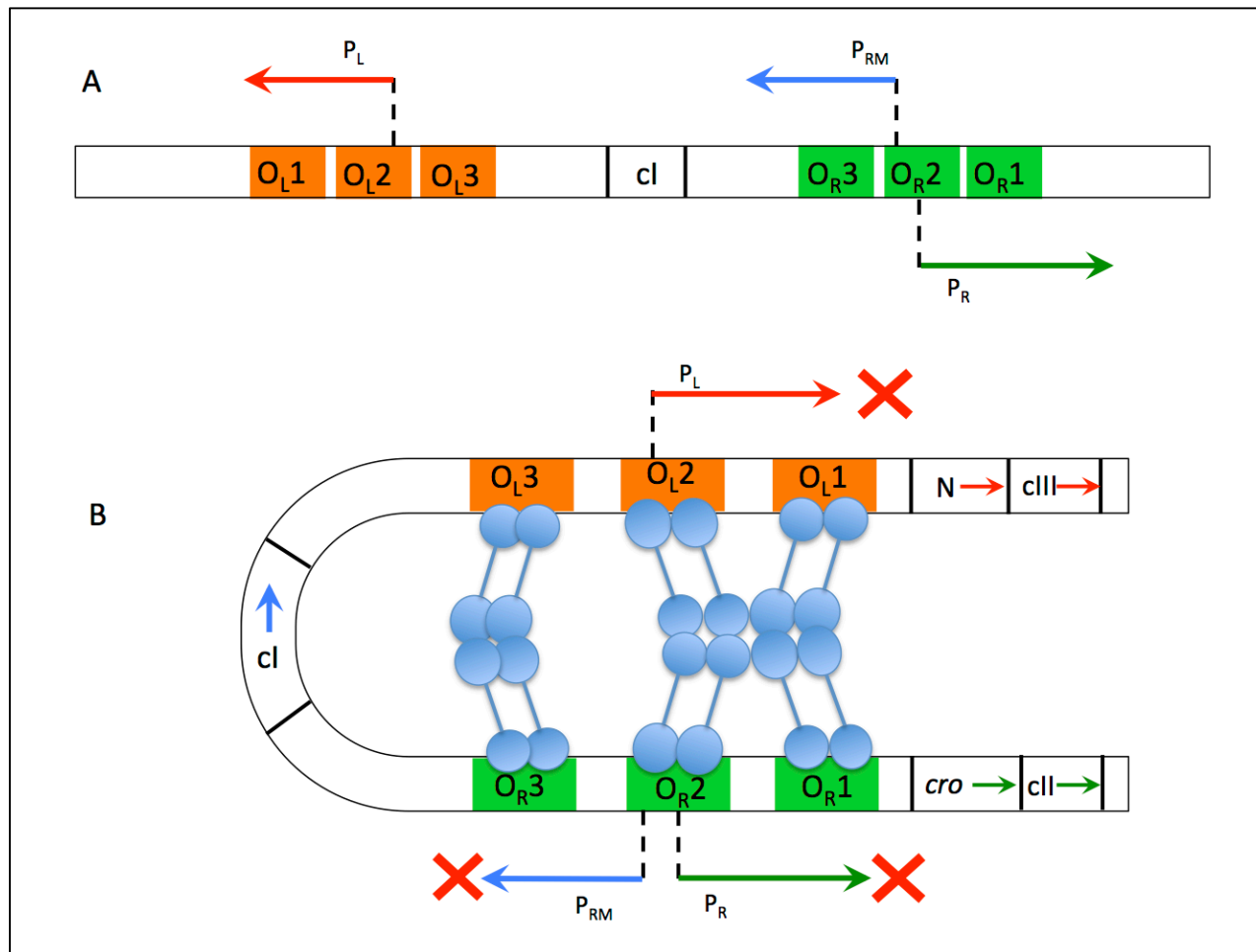


Figure 1.4. The lambda immunity region.

A. The operator O_R is composed of three specific operator binding sites: O_{R1} , O_{R2} and O_{R3} , while O_L composed of operator binding sites: O_{L1} , O_{L2} and O_{L3} . The promoters P_L (red arrow) and P_R (green arrow) control the transcription of the early genes *N* and *cII* and the genes *cro* and *cII*, respectively. The promoter P_{RM} transcribes *cl*.

B. Looping of the operators O_R and O_L takes place after the repressor *CI* binding. The affinity of repressor *CI* for O_{R1} and O_{L1} is greater than its for O_{R2} and O_{L2} . While the affinity of *CI* for O_{R3} and O_{L3} is very low. After its binding to O_{R1} , O_{R2} , O_{L1} and O_{L2} , the levels of *CI* repressor drop in the cell, O_{R3} and O_{L3} become free of repressor, the binding of *CI* to O_{R2} is activates the promoter P_{RM} and the leads to reexpression of *CI* repressor. The new *CI* protein binds to both O_{R3} and O_{L3} blocking the transcription of its own gene (Adapted from Dodd et al., 2001)

However, CI, because it blocks transcription from P_R , prevents the production of Cro. Without the binding of Cro to O_{R3} and O_{R2} , expression of CI can be driven from P_{RM} , which is activated because Cro no longer represses P_{RM} . If CI expression continues, it will eventually self-regulate its own expression by CI binding to O_{R3} and O_L3 in the loop repressing P_{RM} , itself. When levels of CI begin to drop O_{R3} will be unbound enabling re-instigation of CI expression, thus maintaining the lysogenic replication cycle and blocking possible infections by other homoimmune phages (Lewis et al., 2016) (Fig. 1.4).

1.2.3 Lambda phage induction

Once lambda has entered into the lysogenic cycle, the prophage is kept silent by the autoregulation of CI, keeping it stable and allowing the phage genome to be passively replicated with the host genome under the control of host cell replication (Raya and H'bert, 2009; Burnet, 1934). If the lysogen finds itself in conditions that might limit its own survival, the prophage can direct its excision from the bacterial chromosome, through a process called induction and enter into the lytic replication to create new phage particles that will then be released from the host cell by lysis. One of the conditions that the lysogen senses is DNA damage. DNA damage may happen due to exposure to many different factors, e.g. ultraviolet light or certain antibiotics (Yue et al., 2012; Ennis et al., 1985). The accumulation of the ssDNA due to dsDNA damage and breakage turns on about 10-20 of the bacterial survival genes that together are associated with the SOS response (Fonville et al., 2010; Mustard and Little, 2000). The SOS response

involves the activation of RecA, a protein that is ordinarily involved in DNA recombination (Hendrix, 1983) and DNA replication. The RecA protein promotes the autolysis LexA protein, a protein that usually activates the survival genes in *E. coli*. RecA also induces the autolysis of CI protein (Pedersen et al., 2010). This destruction of CI frees O_R and O_L from the repressor, restarting transcription from P_L and P_R (Mustard and Little, 2000).

The reactivation of P_L and P_R is associated with the concomitant expression of Cro and N. CII fails to survive within the cell and two genes downstream of *cIII* are then expressed, the *xis* gene, encoding excisionase (Xis) enzyme, and the *int* gene encoding Int enzyme, under the control of P_L promoter (Court et al., 2007). Integrase now works with its cognate recombination directionality factor (Xis) to drive the excision of the integrated prophage DNA. The lytic cycle starts and the new phages are released (Hendrix, 1983).

There are four genes responsible for cell wall lysis to enable progeny phages to escape the confines of their host cell. These genes are: *S*, encoding a holin; *R*, encoding a transglycosidase, also known as an endolysin; and the *Rz* genes (*Rz* and *Rz1*), encoding an outer membrane lipoprotein which is a type II signal anchor protein (Fig. 1.5). Holins (*S*) are produced as a first step in cell lysis. The *S* protein is responsible for producing non-specific lesions in the inner membrane, IM. These lesions allow the endolysin (*R*) to reach and attack the peptidoglycan, which enzymatically degrades the peptidoglycan (PG) layer of the host

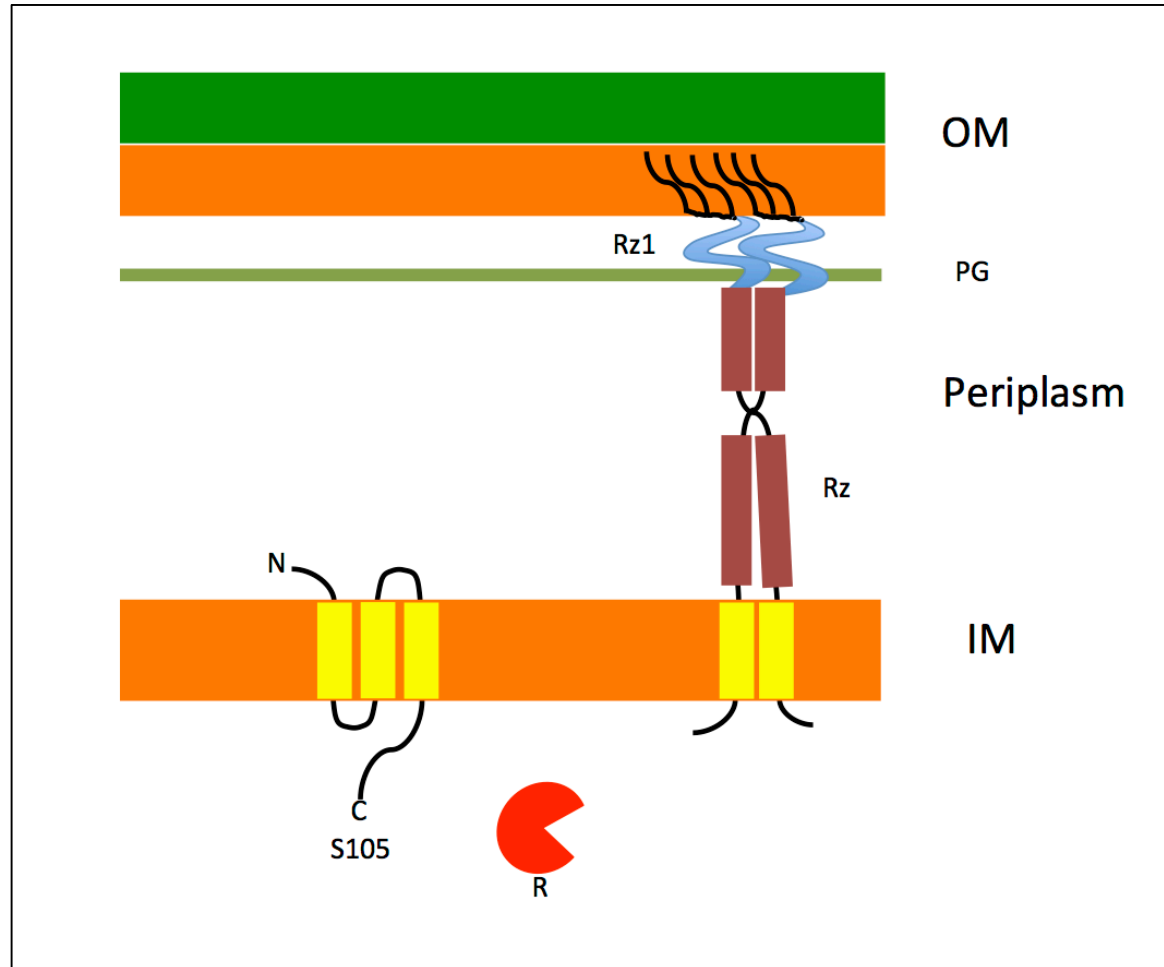


Figure 1.5. Structure and localization of the λ phage Lysis cassette. The endolysin (R) and holin (S105) are localized to the cytoplasm and the inner membrane (IM), respectively. Holin protein forms large lesions in the IM to enable R protein to diffuse into the periplasm space, where its substrate is located, the peptidoglycan (PG). The Rz genes create a spanin complex, the spanin structure composed of two component spanin (Rz = i-spanin and Rz1 = o-spanin). The Rz is embedded into the IM by an N-terminal (yellow) and it has coiled alpha helical periplasmic domain (4 brown lines) separated by linker region. The Rz1 is attached to the outer membrane via the three fatty acyl group (Black color), which is the N-terminal Cys residue. Both of Rz and Rz1 linked by intermolecular disulfide bonds in the periplasmic domains. The Rz/Rz1 spanin is connecting the IM and OM through the meshwork of PG (Rajaure et al., 2015, Berry et al., 2012).

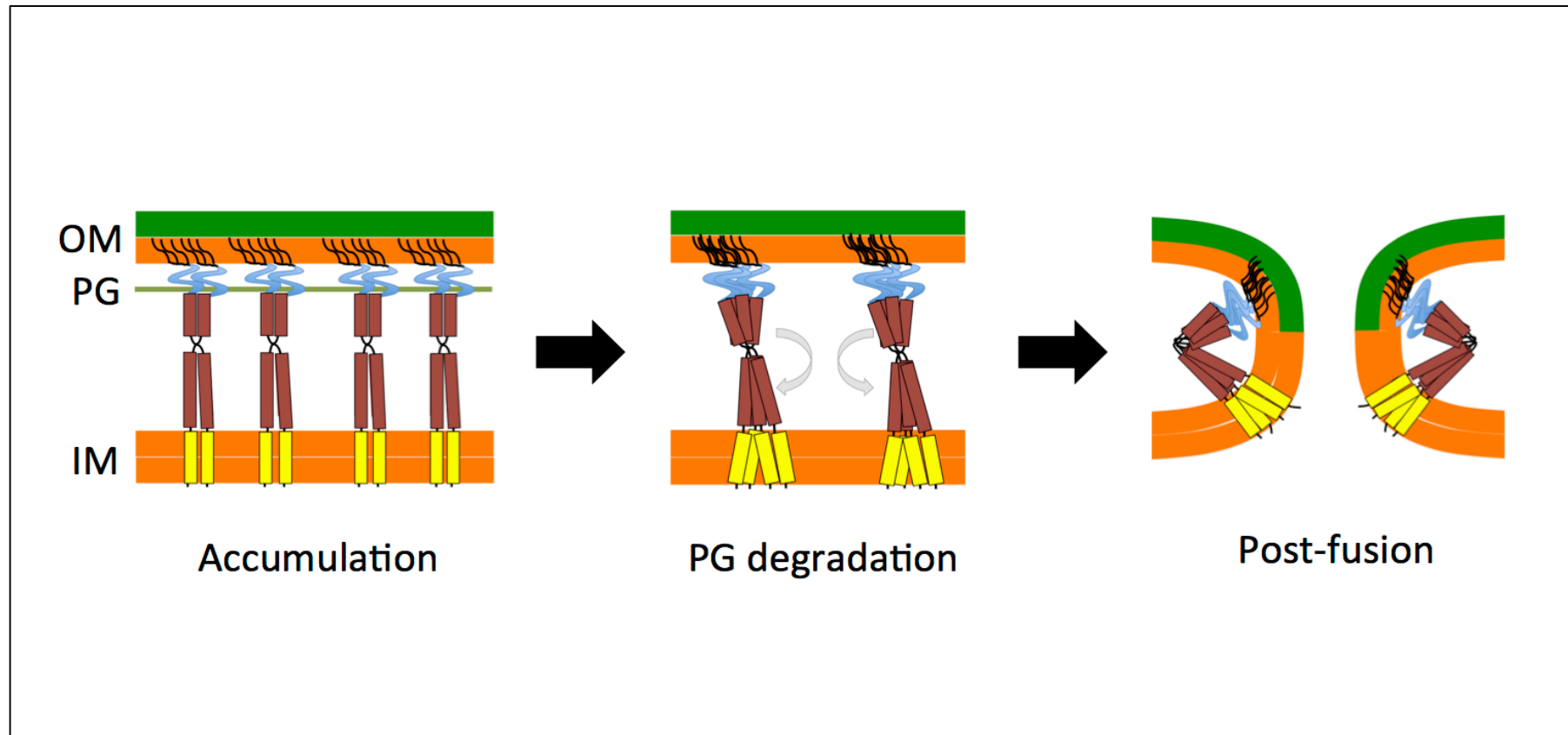


Figure 1.6. The steps in the membrane fusion model for the Rz-Rz1 lytic function. The spanin complexes, connecting the outer membrane (OM) and the inner membrane (IM), accumulate within the lacunae formed by the peptidoglycan cross-linking. Endolysin mediates destruction of the PG liberates the spanins to undergo conformational changes to bring the both IM and OM into contact and stimulate the IM-OM fusion (Adapted from Rajaure et al., 2015).

bacterium (Summer et al., 2007, Young et al., 1979). Rz and Rz1, then form a spanin complex that spans the IM, the outer membrane (OM) and the entire periplasm, enabling fusion of both membranes (Fig.1.6) (Berry et al., 2008).

1.3 Temperate Phage and HGT

Temperate bacteriophages are among the major contributors to the horizontal gene transfer (HGT). HGT involves the movement of DNA between individual cells belonging to: the same bacterial strain or species, cells belonging to the same bacterial family, or even between bacterial cells that are not closely related (de la Cruz and Davies, 2000). In other words, it is the movement of genetic material between bacterial cells other than by the transmission of DNA from parent to offspring, also known as horizontal transmission (Young et al., 1979) Phages can perform this DNA transfer via transduction, or through lysogenic conversion. Using these mechanisms, phages, prophages and phage-like elements play an important role in driving bacterial diversity and bacterial evolution in many ways (Canchaya et al., 2003). They give the infected host immunity against superinfection by the same or related phages. They can work as vectors for the transfer of virulence genes amongst bacterial hosts. In fact a wide variety of bacterial virulence factors are encoded by genes within prophages These factors in some cases may alter the phenotypes of their bacterial host, such as production of an exotoxin or exoenzyme (Holt et al., 2016; Veses-Garcia et al., 2015), or the

expression of modified cell surface protein (Raya and H'bert, 2009; Holt et al., 2016; Riley et al., 2012; Veses-Garcia et al., 2015).

1.3.1 Phage Transduction

Transduction is the movement of genes from a bacterial donor cell to a bacterial recipient cell using phages as the transfer vector. Ikeda and Tomizawa (1965) stated, after their experiments with bacteriophage P1 on *E. coli*, that before the donor host cell lysed following P1 phage induction, the bacterial chromosome had been degraded into small DNA fragments. P1 was able to package any piece of this bacterial DNA, seemingly mistaking it for P1 progeny DNA. This bacterial DNA was then packaged into the phage head in place of phage DNA. When the released progeny phage attack a new *E. coli* cell, the DNA injected by some of the phage particles are actually fragments of DNA from the former bacterial host's chromosome. The newly infected host now becomes the recipient host. The transduced bacterial genes can be incorporated inside the recipient host chromosome through DNA recombination events. Certain phages, like P1 and P22, are able to transduce any gene in the bacterial chromosome, and transduction mediated by these phages is known as *Generalized Transduction* (Stodolsky et al., 1972).

On the other hand, some phages are only able to package a limited number of genes from a donor cell. This type of transduction is called *Specialized*

Transduction. Lambda phage is an example of a phage that is an instrument for specialized transduction. When lambda DNA integrates into the bacterial host chromosome, it does so in a site-specific manner. Lambda Int, a site-specific recombinase, ensures that the lambda genome always integrates at the same site within the *E. coli* genome. When the lambda prophage is induced, Int, with help of Xis and some other DNA binding proteins, ensures that the lambda genome excises precisely from the point of integration to produce the original circular lambda chromosome and to start the lytic cycle (Cho et al., 2002). Very rarely, however, the excision is imprecise and the resultant phage now carries a bacterial gene that lay close to the *attL* or *attR* sites. For λ , the *E. coli* genes close to *attR* and *attL* are the *gal* locus, which encode necessary enzymes for galactose metabolism (Weickert and Adhya, 1993), and *bio*, which encodes biotin protein (Shaw et al., 1999), respectively. The resultant, excised phage genome is defective due to λ genes that are left behind and the host gene that has been acquired. The transducing phages are known as *λ dg_{gal}* (λ -defective *gal*) or *λ db_{bio}* (λ -defective *bio*). The defective chromosomes can be packaged into phage heads, which go on to infect other bacterial hosts. When the specialised transducing phage, *λ dg_{gal}*, infects a new bacterial host, even if it carries a genetically different, resident prophage (Griffiths, 2000), it can insert its genome either by general, homology-dependent recombination or by specific-lambda *att* site, when the *λ dg_{gal}* can thus initiate a new lysogenic pathway, because the defective phage, *λ dg_{gal}*, still encodes the Int enzyme (Griffiths, 1996). The requirements for integrase reaction using the hybrid site *attL* on the *λ dg_{gal}* are different from those for a phage *attP* site (Calendar, 2006). However, *λ dg_{gal}*, in acquiring the *gal* gene will have

left behind some of the tail structural genes (Fig.1.7). Therefore $\lambda dgal$ will be unable to replicate lytically. On the other hand, when the defective $\lambda dbio$, infects another bacterial host, its only able to drive the lytic cycle as it left the *int* gene, or part from it, in previous the host cell genome, but it will still has all lysis, head, and tail genes (Dul and Drexler, 1988; Griffiths, 2000; Shpakovski and Berlin, 1984). This transduction is limited to genes very close to the original integration event and that why it called Specialized Transduction (Fig. 1.7).

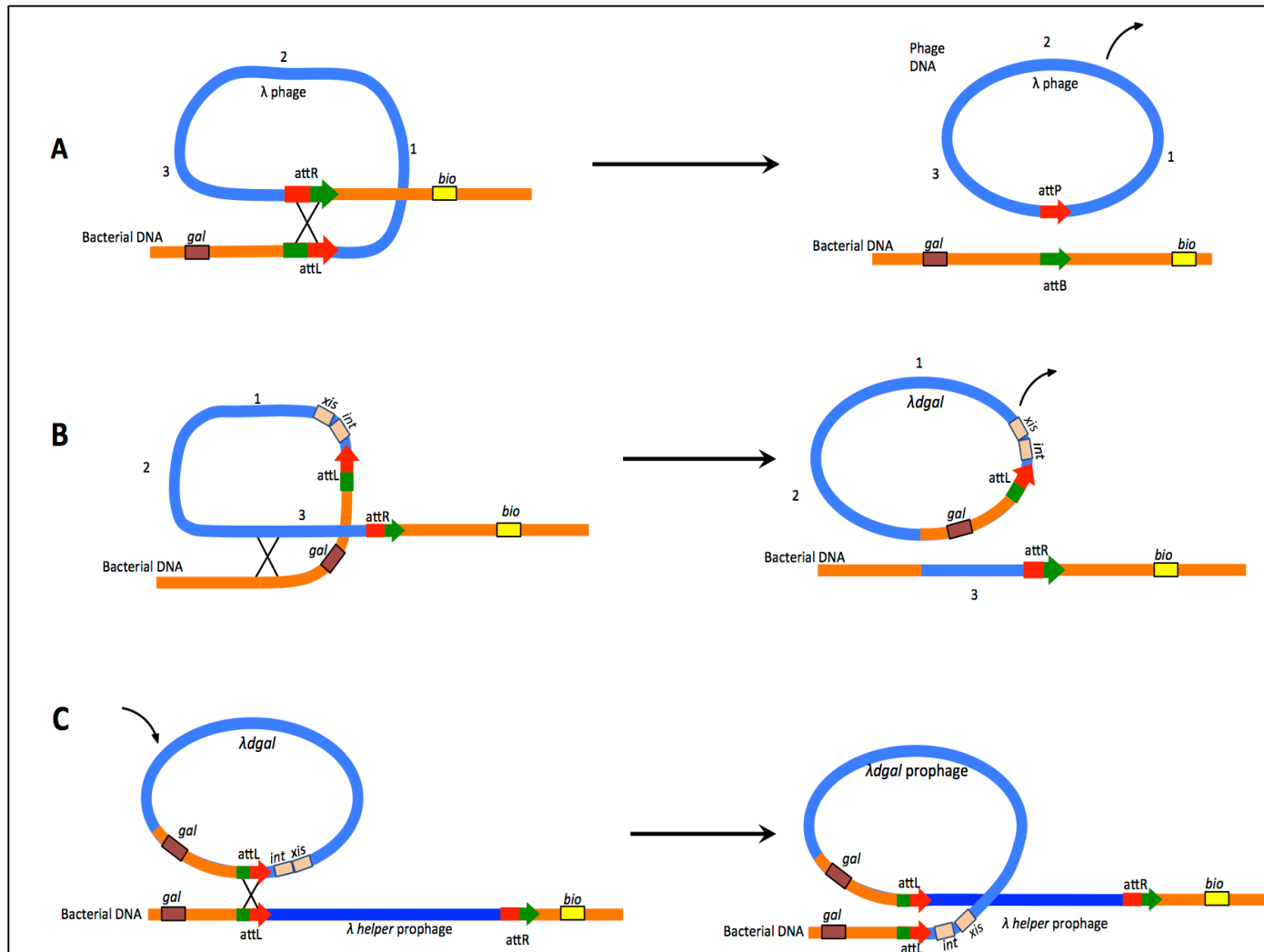


Figure 1.7. Specialized transduction by lambda phage:

A. The normal precise excision. Here *Int* and *Xis* insure that the prophage excises exactly from λ phage *att* sites so that *attR* and *attL* are precisely resolved to restore *attP* and *attB*.

B. In very rare cases, the excision takes place in imprecisely sites. The resulting phage carries the nearby gene (*gal*) and it leaves behind some phage genes. This new phage is called, λ -defective and in this case it is λ -defective *gal* (*λdgal*). Here the phage still encodes *Int* enzyme so it still able to drive site-specific recombination or it can insert its genome inside new host chromosome by general, homology-dependent recombination.

C. When the defective phage infects a new bacterial host with or without a genetically different, resident prophage. The defective phage can still integrate its genome at same specific λ -*att* site, this generate a helping effect by the helper prophage (Adapted from Griffiths, 2000).

1.3.2 Phage Conversion

Phage conversion, or lysogenic conversion, is occurs due to one or more changes in the bacterial host phenotype resulting from infection with a temperate phage. Typically the converting phage becomes a prophage within the chromosome of its bacterial host (Rankin and Platt, 1995), or it becomes an extra chromosomal self replicating unit inside the bacterial cell (Ikeda and Tomizawa, 1968).

Some phages carry genes that are not related to the lytic or lysogenic processes. These genes have the potential to produce new phenotypes for their host cells. They may make the bacterial cell resistant to related phages (McGrath et al., 2002), or increase the pathogenic capacity of the bacterial host cell (now a lysogen) to its host through prophage carriage of new virulence factors (Allison, 2003). In this way, temperate bacteriophages play role in the spread of virulence factors amongst bacteria.

It is also possible for these new bacterial phenotypes to become permanent due to the decay (deletion, insertion and recombination events) of prophage sequences that result in the prophage's loss of their ability to be induced into lytic replication (Hale et al., 2005; Weinbauer, 2004). Many of the bacterial diseases that impact our world are due to bacterial pathogens that were made more pathogenic through the carriage of a converting phage.

1.3.2.1. Down stream effects of phage conversion in *Vibrio cholerae*

The phage that encodes cholera toxin produced by *Vibrio cholerae* is the filamentous temperate phage CTX ϕ . *Vibrio cholerae* acquired the ability to produce cholera toxin only after they became lysogens of this phage (Waldor and Mekalanos, 1996). There are over 200 serotypes of this bacteria species, however only two serotypes, O1 and O139, are causative agents of Asiatic cholera. Asiatic cholera is characterized by watery diarrhea and severe dehydration that can progress rapidly to death (Faruque, 2014). Cholera toxin is solely responsible for the symptoms of Asiatic cholera (Faruque et al., 1999; Waldor and Mekalanos, 1996). Cholera toxin is an oligomeric complex protein (known as AB₅), made of six subunits; five copies of the B subunits, each weighing ~11.6 kDa, producing a ring, surrounding the single copy of the A subunit, which is 28 kDa (Sanchez and Holmgren, 2011). The A subunit is the enzymatic part of the toxin while B subunits bind to specific receptors (GM1 gangliosides) on the cell surface, facilitating the entry of toxic A subunit into the host cell. After the A1 domain of subunit A enters the cytosol, it activates adenylate cyclase drives the cAMP production by G protein stimulation. The cAMP promotes the activation of the cystic fibrosis transmembrane conductance regulator (CFTR), leading to watery diarrhea (Bharati and Ganguly, 2011).

1.3.2.2. Down stream effects of phage conversion in *Corynebacterium diphtheriae*

Diphtheria toxin that produced by *Corynebacterium diphtheriae* is actually encoded by a temperate corynephage genome that carries the *tox* gene. *Corynebacterium diphtheriae* acquired the ability to produce diphtheria toxin only after they become lysogens of this phage (Groman, 1984).

Diphtheria toxin is a 60 kDa molecular weight protein composed of two subunits A and B. (Groman, 1984). Subunit B facilitates the entry of toxin subunit A into the human host cell by binding to the heparin-binding, EGF-like growth factor receptor on the cell surface. The A subunit enters the cell within an endosome via receptor-mediated endocytosis. The B subunit creates pores in the endosome membrane, allowing the escape of the A subunit into the cell cytoplasm. The A subunit inhibits the synthesis of new protein in the affected cell (Collier, 1975).

The disease diphtheria is an acute infection, characterised most often by a sore throat, fever (38 °C or above), and the development of an adherent membrane on the tonsils and/or nasopharynx called a pseudomembrane. This pseudomembrane is usually produced at the primary site of the infection, normally the mucous membranes of the respiratory tract. From these pseudomembranes diphtheria toxin can spread throughout the body to cause systemic toxic infections (Waldor and Mekalanos, 1996).

1.3.2.3. Shiga toxin-producing and enterohaemorrhagic *Escherichia coli*

The bacteriophage, $\Phi 24_B$, is a lambdoid phage (Stx phage) that gives the enterohaemorrhagic *E. coli* (EHEC) their ability to produce Shiga toxin (Stx), only after they become lysogens of these phages. (Paton and Paton, 1998). The Enterohaemorrhagic *E. coli* (EHEC) is a subset of pathogenic *E. coli* that can cause diarrhea or hemorrhagic colitis in humans. Hemorrhagic colitis, caused by Enterohaemorrhagic *E. coli* (EHEC), occasionally progresses to hemolytic uremic syndrome (HUS), an important cause of acute renal failure in children and morbidity and mortality in adults (Geue et al., 2009; Xu et al., 2012).

The main reservoirs of EHEC serotypes associated with human infections are ruminants, in particular cattle (Geue et al., 2009; Xu et al., 2012). They can be transmitted to humans through the consumption of contaminated food and/or drink, and even person-to-person spread has been reported (Nataro and Kaper, 1998). One of the most important factors contributing to the pathogenesis of EHEC infection, and the single most important factor in the severe symptoms, is the production of Shiga toxin (Stx) (Paton and Paton, 1998). Stx is encoded by two small genes (*stxA* and *stxB*), which are arranged in an operon structure.

Stx belongs to a family of AB₅ toxins comprising of a single A subunit that is associated with pentameric B subunits that are responsible for toxin binding to its

receptors in bacterial cells (Lingwood, 1999). The A subunit is an N-glycosidase which cleaves the N-glycosidic bond of specific adenine residues within rRNA, inactivating the 60S ribosomal subunits leading to protein synthesis inhibition. Moreover, it has been stated that Shiga toxin induce apoptosis in many cell types (Sandvig, 2001; Law, 2000).

Stx production is directly responsible for two dangerous EHEC infection sequelae, Hemolytic Uremic Syndrome (HUS) and Thrombotic Thrombocytopenic Purpura (TTP) (Paton and Paton, 1998; Paton and Paton, 1999). Although the global incidence of EHEC infection is low, there is an unacceptably high proportion of infected individuals that go on to develop severe disease (10%) and death (3-5%) (Watanabe et al., 1999). The *stx* genes are acquired by the bacterial cell following infection by a bacteriophage carrying the *stx* genes, like Stx phage $\Phi 24_B$. The expression of the toxin genes is then controlled by the bacteriophage's lytic replication cycle, which subverts the host cell's resources in order to produce more phages in the lytic cycle (Bradley, 1967).

1.3.2.4 Converting phage summary

In summary, phages can move genes from one bacterial host cell to another new cell through the process of phage conversion. Phage conversion can result in bacterial cells gaining abilities that they did not possess prior to phage infection. Additionally, once the phage DNA is in the cell it is free to recombine with cellular

DNA sequences to produce genetic sequences new to both phage and bacterial host.

These biological events play an efficient role in driving the diversification and evolution of the bacterial cells (Allison 2003). The reasons behind the ability of phages to mediate these biological events are: their capacity to inject their genome inside the host cell's cytoplasm to become: 1) an extra chromosomal self-replication unit (Ikeda and Tomizawa, 1968), or 2) integrate their genome inside the host cell's chromosome and become a prophage (Groman, 1984). For the incoming phage to become a prophage, a phage encoded recombinase enzyme must be produced to direct prophage integration into the bacterial chromosome. These enzymes in general are called site-specific recombinases. Phage site-specific recombinases, in the absence of any other phage-encoded factors catalyze the site-specific, uni-directional recombination reaction (Serra-Moreno et al., 2008).

1.4 Site-specific recombination

Site-specific DNA recombination is a genetic recombination process in which DNA strand exchange take place between two DNA molecules that share DNA sequence homology (Kilby et al., 1993). Site-specific recombinases promote DNA fragment rearrangements by identifying and binding to a short DNA sequence called *att*

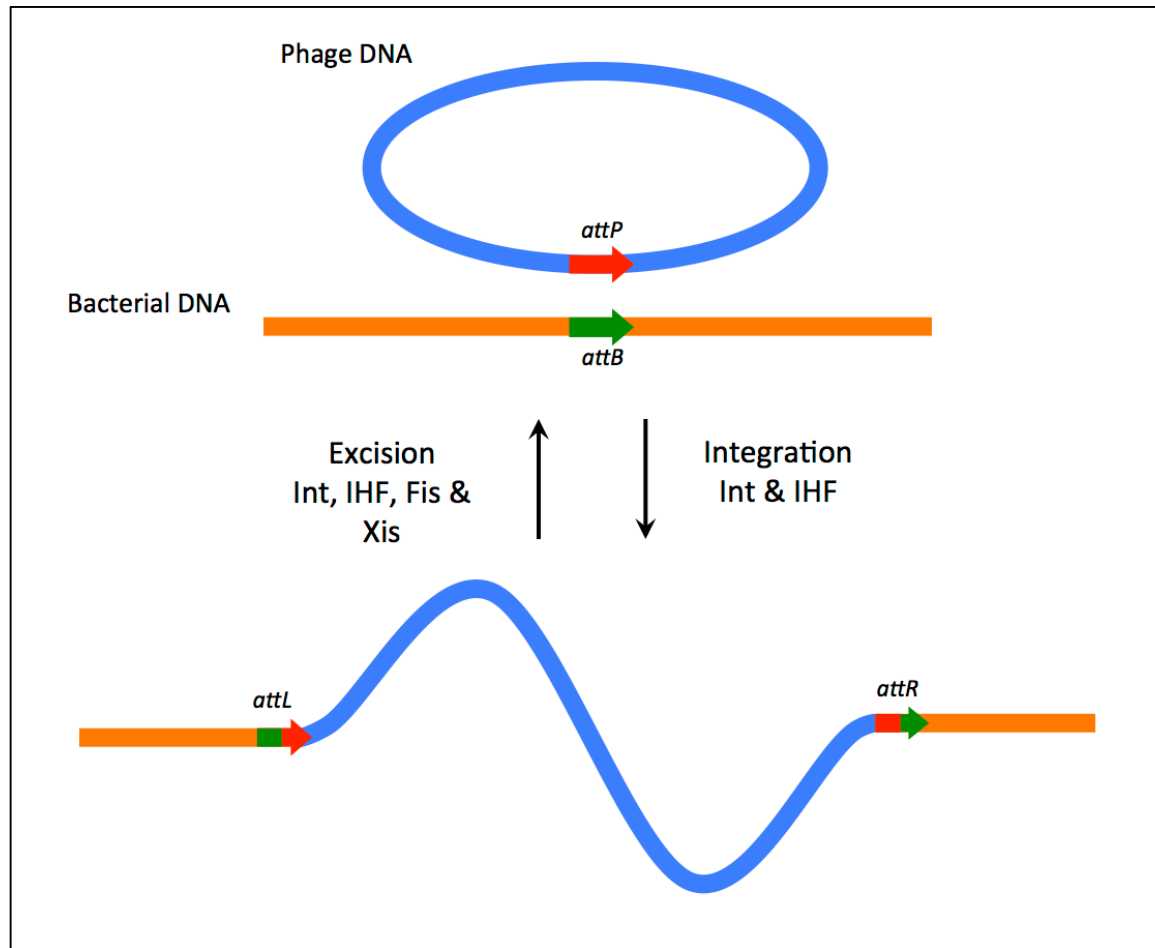


Figure 1.8. Lambda phage site-specific integration and excision. After injection of the lambda genome into the host cytoplasm, the phage genome is circularised (Blue circle). The phage DNA then, integrates itself via the phage attachment site (*attP*; Red arrow), into the host or bacterial attachment site (*attB*; Green arrow). This reaction creates the prophage flanked by the new hybrid attachment sites *attR* and *attL*. These sites are the start points of the excision reaction to regenerate *attP* on the excised phage DNA and *attB* on bacterial chromosome. The phage encoded integrase enzyme (Int) and the host-encoded integration host factor (IHF, a DNA binding protein that stabilizes the DNA to support the action of Int) are both required for integration and excision event. However, excision requires two additional proteins: Xis and Fis. Xis is a phage encoded recombination directionality factor, while Fis is Factor for Inversion Stimulation, encoded by *E coli* *fis* gene (Adapted from Groth and Calos, 2004).

sites, at which these enzymes cleave the DNA molecules, and direct the exchange between the two DNA molecules (Fig. 1.8)(Craig, 1988).

Site-specific recombination approaches are highly specific, fast and efficient, even in some complex eukaryotic genomes (Sauer, 1998). Site-specific recombination systems are involved in different types of cellular processes, including but not exclusive to bacterial pathogenesis, bacterial genome replication and the movement of mobile genetic elements around a single bacterial cell's genome. These abilities make these systems useful for the development of genetic engineering tools (Kilby et al., 1993).

Some site-specific recombination systems, like that derived by ϕ C31, need nothing more than the recombinase enzyme and recombination sites to function (Thorpe and Smith, 1998). Other systems, like that derived by λ phage, require a number of accessory proteins to function. These accessory proteins are either encoded by the bacterial host or are phage encoded. Site-specific recombination does not need any external energy source. The reaction takes place and ends without the reaction takes place and ends without loss or gain of any part of DNA substrates (Fogg et al., 2014).

The substrates that a temperate phage's site-specific recombinase needs to drive integration into the bacterial host chromosome are DNA sequences, typically between 30-250bp, within the phage genome called *attP* (phage attachment site) and a DNA sequence within the host chromosome named *attB* (bacterial attachment site). The recombination reaction between these two sites creates an integrated prophage flanked by the new attachment sites *attR* and *attL*. These sites, *attL* and *attR*, are hybrid sites of *attP* and *attB* that can later serve as the start points for the excision reaction to regenerate *attP* on the excised phage DNA and *attB* within the bacterial chromosome. Int is the main enzyme in both of these recombination reactions, and it is the only phage-encoded protein that is required to facilitate the integration reaction. Excision also requires a phage encoded Xis or recombination directionality factor (RDF) (Yin et al., 1985).

Lambda site-specific recombinase or integrase (Int) functions as a homotetramer, recognizing specific recombination sites (*attB* and *attP*) in both the bacterial and phage DNA, respectively. The Lambda bacterial attachment site, *attB*, is 21 bp while its phage attachment site, *attP*, is ~ 240 bp. These two sites share an overlap region of only 7 bp, which serves as the core for recombination. The tetrameric integrase binds the two DNA substrates to form an intermediate Holliday junction, which is resolved to form new hybrid recombination sites *attL* and *attR* (Fig. 1.8). Lambda integrase cannot drive the DNA recombination without the presence of the host encoded accessory protein Integration Host Factor (IHF) (Craig and Nash, 1984).

The new hybrid sites *attL* and *attR* are the start points of the excision reaction to regenerate *attP* on the excised phage DNA and *attB* on bacterial chromosome (Cassell et al., 1999).

Site-specific recombinases can be divided according to amino acid sequence homology and mechanism of action into two main families, tyrosine and serine recombinases (Table 1.1). Both of these families are populated by recombinases that drive a variety of recombination functions. These functions include: phage integration and excision, resolution of DNA concatemers (Ritchie and Joicey, 1978) and DNA inversion events associated with phase and antigenic variation mechanisms at contingency loci (Robles-Oteiza et al., 2015).

1.4.1. Serine integrases

The serine family of site-specific recombinases use a serine residue as the catalytic residue to drive DNA recombination. This family is populated by enzymes like the invertase Gin (Kahmann et al., 1984), the resolvases from Tn3 (Krasnow and Cozzarelli, 1983) and $\gamma\delta$ (Reed et al., 1982) and the phage recombinases from ϕ C31 (Kuhstoss and Rao, 1991), R4 (Groth et al., 2000), TP901-1 (Christiansen et al., 1996) and ϕ FC1 (Yang et al., 2002). The function of the latter four phage encoded recombinases have been studied in depth, and can be used as representatives of most of the family members (Table 1.1).

The *att* sites that are targeted by serine recombinases are typically ~50 bp, composed of a variety of inverted repeat sequences that are flanked by a short central region of 2-12 bp at the crossover core (Groth et al., 2000).

Serine site-specific recombinases bind to their specific attachment sites, however only two subunits (dimers) are active to drive the DNA exchanges (Fig.1.7), as opposed to the active tetramers of integrase (described in 1.4.3 Tyrosine recombinases). The synaptic complex is where the catalytic serine residues act to break all four DNA strands at the same time (Groth et al., 2000). The double stranded DNA is cleaved at the 3' end of two nucleotides overhangs followed by the formation of phosphoseryl covalent bonds between the four free 5' ends and each integrase monomer. An integrase subunit rotation process takes place when two of the enzyme subunits, that remain covalently bound to the DNA 5' ends, rotate 180° (Fig.1.9). This process swaps the two half sites one instead of the other and creates new recombination products (Fig. 1.9). The core bases have to be same in both parent attachments sites for the integrase to drive recombination (Ghosh et al., 2003).

Table 1.1: Selected, well characterised site-specific recombinases and their requirements for recombination. None of the Stx phage integrases or their requirements have been well characterized, yet.

(Adapted from Fogg et al., 2014 & Groth and Calos, 2004)

| Site-specific recombinase family | Phage name | Host Genera | attB (bp) | attP (bp) | Required Host factor | RDF | References |
|----------------------------------|------------|----------------------|-----------|-----------|----------------------|-------------|--|
| Serine | ϕ C31 | <i>Streptomyces</i> | 34 | 39 | | gp3 | (Groth et al., 2000, Kuhstoss and Rao, 1991) |
| | ϕ BT1 | <i>Streptomyces</i> | 36 | 48 | | gp3 | (Zhang et al., 2008b) |
| | TG1 | <i>Streptomyces</i> | 39 | 43 | | gp25 | (Morita et al., 2009) |
| | Bxb1 | <i>Mycobacterium</i> | 38 | 48 | | gp47 | (Ghosh et al., 2006) |
| | ϕ Rv1 | <i>Mycobacterium</i> | 40 | 52 | | Rv1584c/Xis | (Bibb et al., 2005) |
| | TP901-1 | <i>Lactococcus</i> | 31 | 50 | | ORF7 | (Breüner et al., 1999, Christiansen et al., 1996) |
| Tyrosine | λ | <i>Escherichia</i> | 21 | 240 | IHF | Xis | (Abremski and Gottesman, 1982, Enquist et al., 1979) |
| | L5 | <i>Mycobacterium</i> | 29 | 240 | mIHF | Gp36/Xis | (Lee and Hatfull, 1993, Lewis and Hatfull, 2000) |
| | P2 | <i>Salmonella</i> | 17 | 220 | IHF | Cox | (Yu and Haggård-Ljungquist, 1993) |
| | P22 | <i>Salmonella</i> | 27 | 260 | IHF | Xis | (Mattis et al., 2008, Leong et al., 1985) |
| | HP1 | <i>Hemophilus</i> | 18 | 420 | IHF | Cox | (Waldman et al., 1986, Hauser and Scocca, 1992) |

In the phage ϕ C31 the crossover core region is TTG, flanked by inverted repeat sequences. The repeats in the *attP* site are a bit farther from the core than in *attB* (Thorpe et al., 2000). ϕ C31 serine site specific recombinase performs DNA recombination between bp 39 in the *attP* site and bp 34 in the *attB* site (Groth et al., 2000). Similar simple *att* sites can be also found in Phages R4 and TP901-1 and their target hosts, all utilising minimal sites of 56 bp or less for both *attB* and *attP* (Olivares et al., 2001). It has been reported that for efficient recombination in both ϕ C31 and TP901-1, 2bp in the core region have to be conserved; TT for ϕ C31 in the TTG core region (Combes et al., 2002) and TC in the TP901-1 TCAAT core region (Breüner et al., 2001).

Unlike most tyrosine integrases, serine integrases can mediate the integration of their phage genome inside the bacterial host chromosome without the requirement of other accessory factors (Fogg et al., 2014). Serine integrases are inactive on *attR* and *attL*; however, the excision reaction has been observed in some cases (Breüner et al., 1999). For efficient excision to occur, the serine integrases require the action of a phage-encoded recombination directionality factor, RDF. The recombination directionality factor is a protein that drives phage excision and functions in one of two ways: by creating a synapsis between the *attL* and *attR* sites, activating excision or by reducing integration by inactivating the synapse between the *attB* and *attP* sites (Bibb et al., 2005).

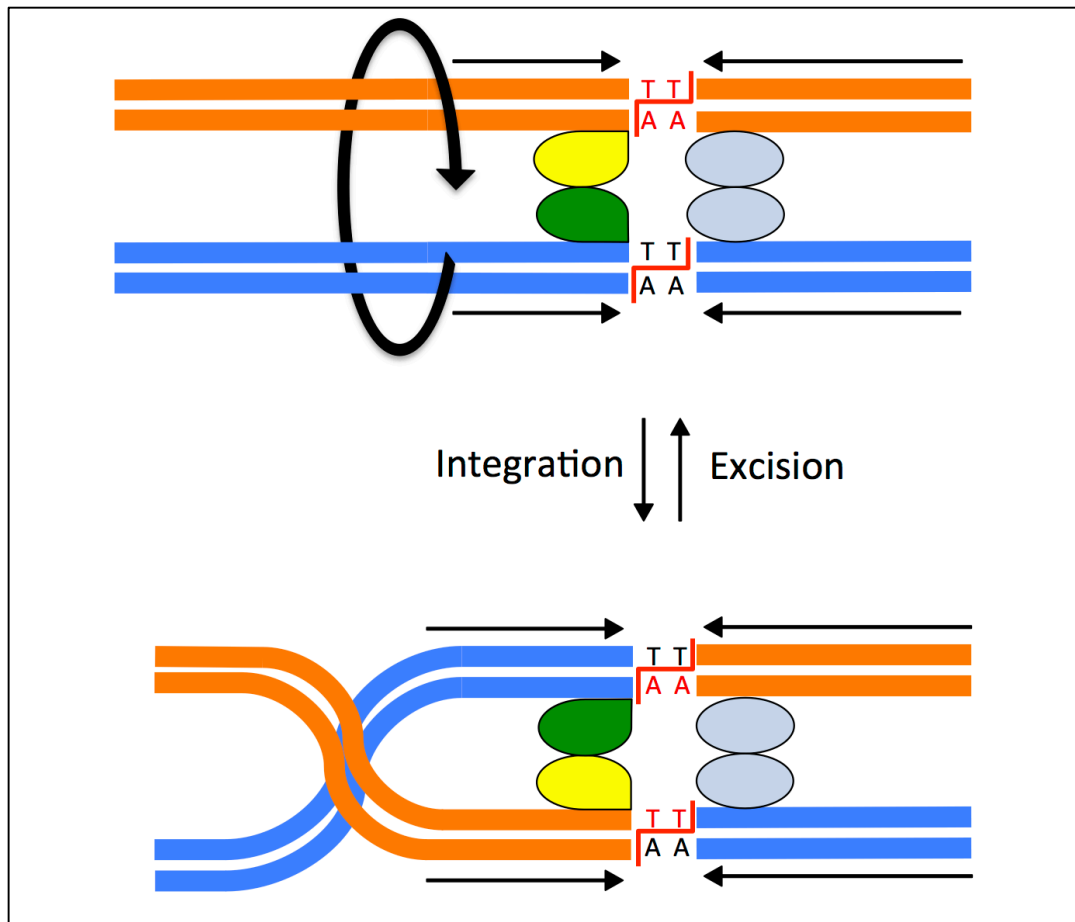


Figure 1.9. Overview of the cleavage and rejoining of attachment sites by a serine recombinase. The thick blue and orange lines represent the DNA strands with central dinucleotide shown, in this case, two T:A base pairs . the red lines represent the position of the breaks made when the recombinase cuts the DNA strands. The subunit rotation process exchanges the DNA half sites when two subunits of the recombinase (highlighted by green and yellow colors), bound to the half sites from opposing DNA strands, rotate 180° compared the other pair of subunits which stay waiting, to rejoin the DNA strands again in a new recombination products (Adapted from Fogg et al., 2014).

1.4.2 Applications of serine site-specific recombinases

The minimal requirements of serine site-specific recombinases for both the DNA substrates and the host encoded accessory factors, make many *in vitro* and *in situ* DNA assembly strategies far easier to design and optimize compared to standard cloning strategies. This is particularly apparent when serine recombination systems are used in heterologous cells that may share very little of the molecular machinery found in the serine site-specific recombinase's natural host (Fogg et al., 2014).

Among serine recombinases, the activity of ϕ C31 recombinase was exploited in the development of integration vectors for use in its natural host, *Streptomyces* as it contains the ϕ C31 *attB* site in its genome (Kuhstoss and Rao, 1991). Due to its high efficiency and extreme stability these plasmids become efficient vectors for introducing DNA sequences into the chromosome of *Streptomyces spp.* and related bacteria, and the ϕ C31 enzyme can be used for accurate knock-out of target DNA sequences (Zhang et al., 2013). In fact, whole biosynthesis clusters of antibiotic resistance genes have been integrated into heterologous host genomes using ϕ C31 recombinase (Zhang et al., 2013).

The ϕ C31 recombinase activity has also been exploited to improve applications that utilize Cre and FLP systems in sequences exchange and the engineering of heterologous cells. It has been reported that an *attB*-flanked cassette can be

integrated, using the activity of the ϕ C31 recombinase inside mouse embryonic stem cells (ES cells) in *attP*-flanked cassette in their genome. This has been performed by placing the DNA sequence flanked by either *attP* or *attB* sites into genome of ES cells (Docking site). Then, a plasmid containing a sequence flanked either by *attP* or *attB* sites (Incoming sequence) has been introduced in the ES cells (Belteki et al., 2003).

Similar study has demonstrated the ability of the same recombinase to introduce a DNA sequence flanked by two *attB* sites inside cells of the yeast *Schizosaccharomyces pombe* into an already chromosomally inserted *attP* site. The controlled directionality of ϕ C31 recombinase makes it much better suited for cassette exchange applications than Cre or FLP tyrosine integrases (Thomason et al., 2001).

An expanded potential for the ϕ C31 recombinase is coming from its competence to recognize mouse and human derived pseudo *attP* sites (genetically inserted *attP* site), and use them to drive integration. An *attB* harboring plasmid can be integrated into human cells within a genomic pseudo *attP* site with a very good integration frequency and holds much promise in the applications of gene therapy (Thyagarajan et al., 2001). The integrases R4 and TP901-1 are also members of the serine recombinase family and are being successfully utilized in other mammalian cells applications (Olivares et al., 2001; Stoll et al., 2002).

1.4.3 Tyrosine recombinases

Members of the tyrosine site-specific recombinase family, also generally referred to as integrases, use the amino acid tyrosine within their active site to attack the DNA backbone during cleavage. This family is currently composed of more than 300 model enzymes (Groth and Calos, 2004). The recombination mechanisms used by Int from λ and Cre from phage P1 are representatives of almost all tyrosine integrases (Table 1.1)) (Austin et al., 1981; Campbell, 1962)

In most cases, the *attB* site recognised by tyrosine recombinases is composed of two inverted repeats flanking the core region. The inverted repeats are called core-type binding sites and range from 9 bp to 13 bp. The core or the overlap, or some times called the crossover region, is usually 6 bp – 8 bp. The site *attP* is more complicated, containing several adjacent DNA binding sites called arm-type sites (this going to be discussed in Chapter 3), which are direct repeats that flank the core-type sites (Ross et al., 1979). Tyrosine integrases bind to the core-type binding sites of *attP* and *attB* and bring the two sites together in a tetrameric synapse (Fig.1.10). The creation of the correct complex or synapse is essential for activation of the recombination activity of the integrases (Biswas et al., 2005). The catalytic residues (tyrosine) for this enzyme in two of its monomers within the tetramer, grab both *attP* and *attB* together to create a recombination joint called a Holliday junction. This junction is then resolved by the integrase to form new, hybrid recombination sites *attR* and *attL* (this also going to be discussed in

Chapter 3). The only phage encoded protein needed for DNA recombination is Int. However, in order to drive its action, Int requires the host encoded integration host factor, IHF, which binds to the *attP* arms and introduces sharp bends that are necessary for the creation of an integration competent synaptic complex (Craig and Nash, 1984) (Fig.1.10).

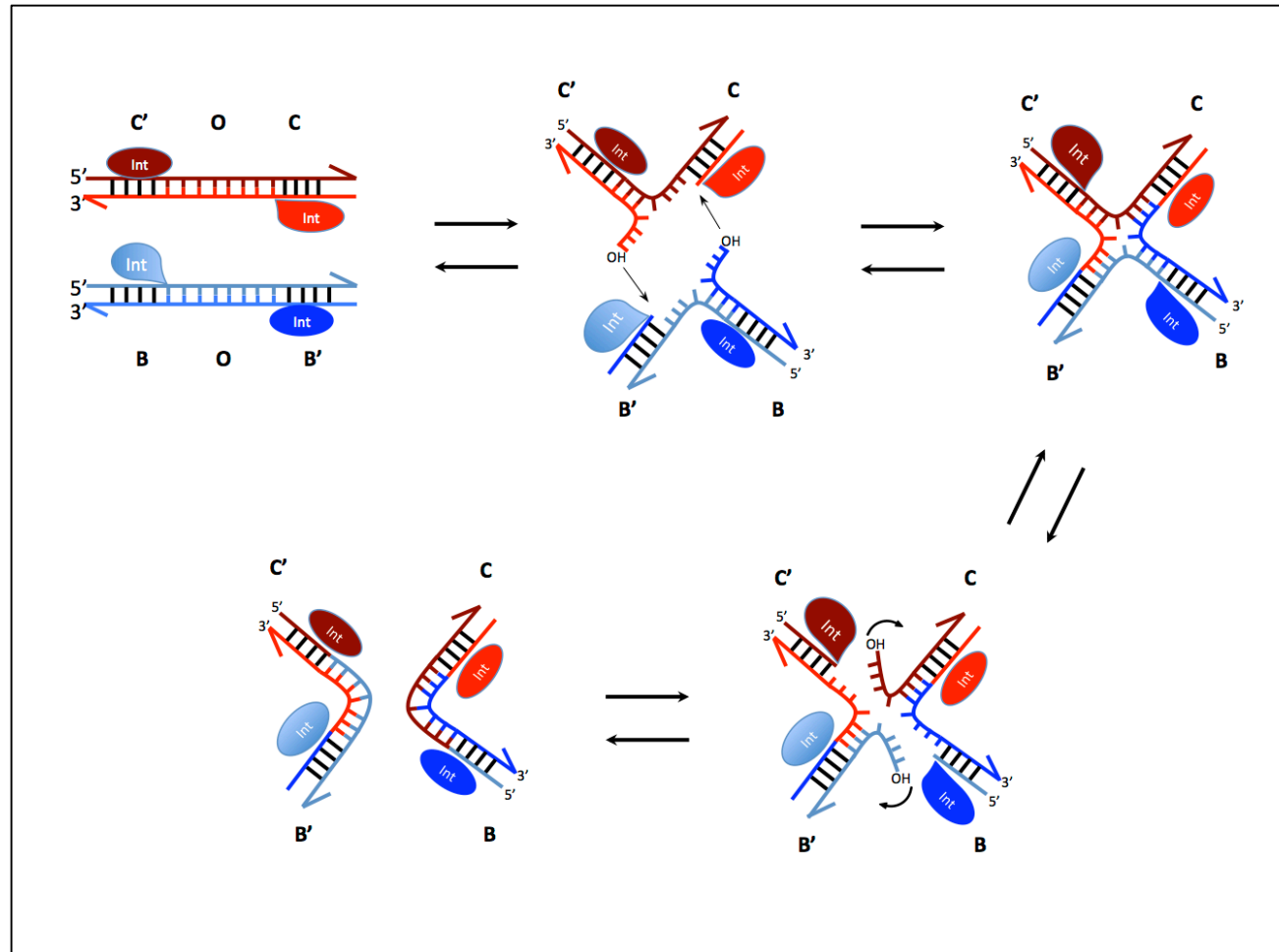


Figure 1.10. DNA recombination by lambda integrase and most tyrosine recombinases taken from (Laxmikanthan et al., 2016). Four protein monomers bind to the DNA strands. The reaction starts with two active integrase monomers performing the first set of DNA cleavages in both attB and attP sites and two inactive monomers. The result of the first cleavage set is the formation of a 3'-phosphotyrosyl intermediate and free 5' hydroxyl groups and strands free 5' hydroxyl groups and strands exchanges to create a Holliday junction (HJ) intermediate. Once the HJ intermediate is formed, the second pair of integrase monomers becomes active to perform the second set of strands exchanges resulting in the final ligation of the recombination mechanism. The host-encoded integration host factor (IHF, a DNA binding protein that stabilizes the DNA conformationally to support the action of Int) is necessary for recombination derived by tyrosine recombinase (Adapted from Rajeev et al., 2009).

The newly formed *attR* and *attL* sites, created from the fusion of the *attP* and *attB* sites, are not substrates for Int or IHF binding. Resolution of the *attL* and *attR* sites to excise the now integrated prophage requires an additional phage-encoded recombination directionality factor (RDF, also known as excisionase [Xis]) in addition to integrase and IHF, and the host encoded protein Fis (Factor for Inversion Stimulation, encoded by *E. coli* *fis* gene, Gene ID: 947697). The four proteins work together in binding to the P-arm in *attR* inducing a loop shape that rearranges the tetrameric integrase complex bound to the core sequence. Sequential DNA cleavage and strand exchange reaction to reproduce *attP* and *attB* sites (Fig. 1.10)(Lewis and Hatfull, 2001).

In addition to the catalytic tyrosine (T342 in λ), five other amino acids are highly conserved in the tyrosine family of recombinases, three of them corresponding to λ R212, H308, and R311, which all together comprise what is known as the RHR triad. The two other amino acids are lysine and histidine residues, which are present in >90% of the family members, corresponding to λ , K235 and H333, respectively, and making the classic pentad RKHRH (Nunes-Düby et al., 1998). Lambda integrase is composed of 356 amino acids and possesses two functional domains. The N-terminal domain, comprised of residues 1-64, is responsible for binding to the arm type binding sites of *attP*. The C-terminal domain, which contains the catalytic residues, binds at lower affinity to the core-type sites. The

action of the C-terminus can be separated into two activities; amino acids that are responsible for the core type binding, residues 65-169, and those associated with the catalytic domain, residues 170-356. The catalytic tyrosine and the RKHRH pentad are located in the latter region (Tirumalai et al., 1998) (Fig. 1.11). The crystal structure, solved by Kwon *et al*, of the catalytic C-terminal domain showed that the RKHRH pentad residues are located close to each other in a groove expected to be the active site. Both N-terminal and C-terminal are participating in the protein-protein interactions between the integrase monomers (Kwon et al., 1997).

The interaction between tyrosine recombinases and their *att* sites are highly specific. λ present Int is so closely related to HK022 Int. They share about 70 % homology and both integrases can recognise same *attP* arm-type binding sites. Nevertheless, neither integrase can recombine the *att* sites of the other phage. The particular DNA base pairs that are responsible for this specificity are present in the core type-sites.

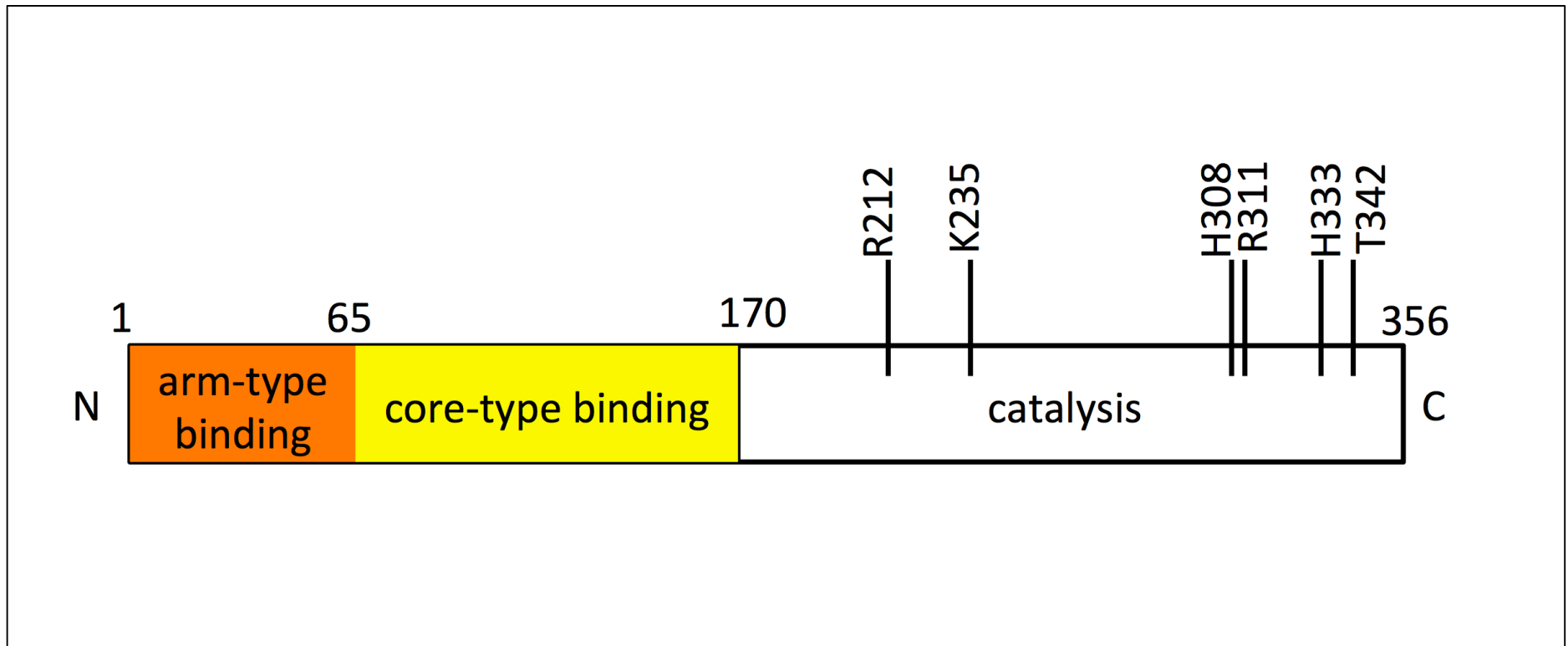


Figure 1.11. Protein domain structure of λ Int. The orange and yellow coloured regions are the DNA binding domains, which are separated by the arm type and the core type binding domains. The white region is catalytic domain. The conserved catalytic residues are labelled. The catalytic domain is the residues 170-356. In addition to the catalytic tyrosine residue (T342 in λ), five other amino acids are highly conserved in the tyrosine family of recombinases; three of them λ R212, H308, and R311, together comprise what is known as the RHR triad. The other common amino acids (λ K235 and H333) are present in >90% of the family members (Adapted from Groth and Calos, 2004).

These sites are B' sites in λ and HK022, the site C in HK022 and some close bases in the flanking sequences of these cores sites (Nagaraja and Weisberg, 1990).

1.4.4 Applications of tyrosine recombinases

Due to their ability to specifically and efficiently recombine DNA sequences in living cells, tyrosine recombinases, or integrases, underpin many biotechnology techniques.

1.4.4.1 Gateway® cloning system

The Gateway® cloning method is the most widely used technique reliant on tyrosine recombinases. Gateway® is an *in vitro* application that uses λ integrase to facilitate fast cloning of linear DNA into especially designed plasmid vectors. Once the desired sequence has been cloned it can be moved with ease into various other specific plasmids for downstream use (*e.g.* induction vectors for protein expression or protein purification). This system uses the ability of the λ integrase to promote integration and excision reactions with high accuracy and specificity to rearrange *att* sites that flank the multiple cloning sites of the Gateway® vectors (Freuler et al., 2008).

The Gateway® systems works when the DNA insert of interest is cloned using cassette exchange into a Gateway® entry plasmid. The entry plasmid is designed so that the cloned gene of interest is flanked by *attL1* and *attL2* after *attB/attP*

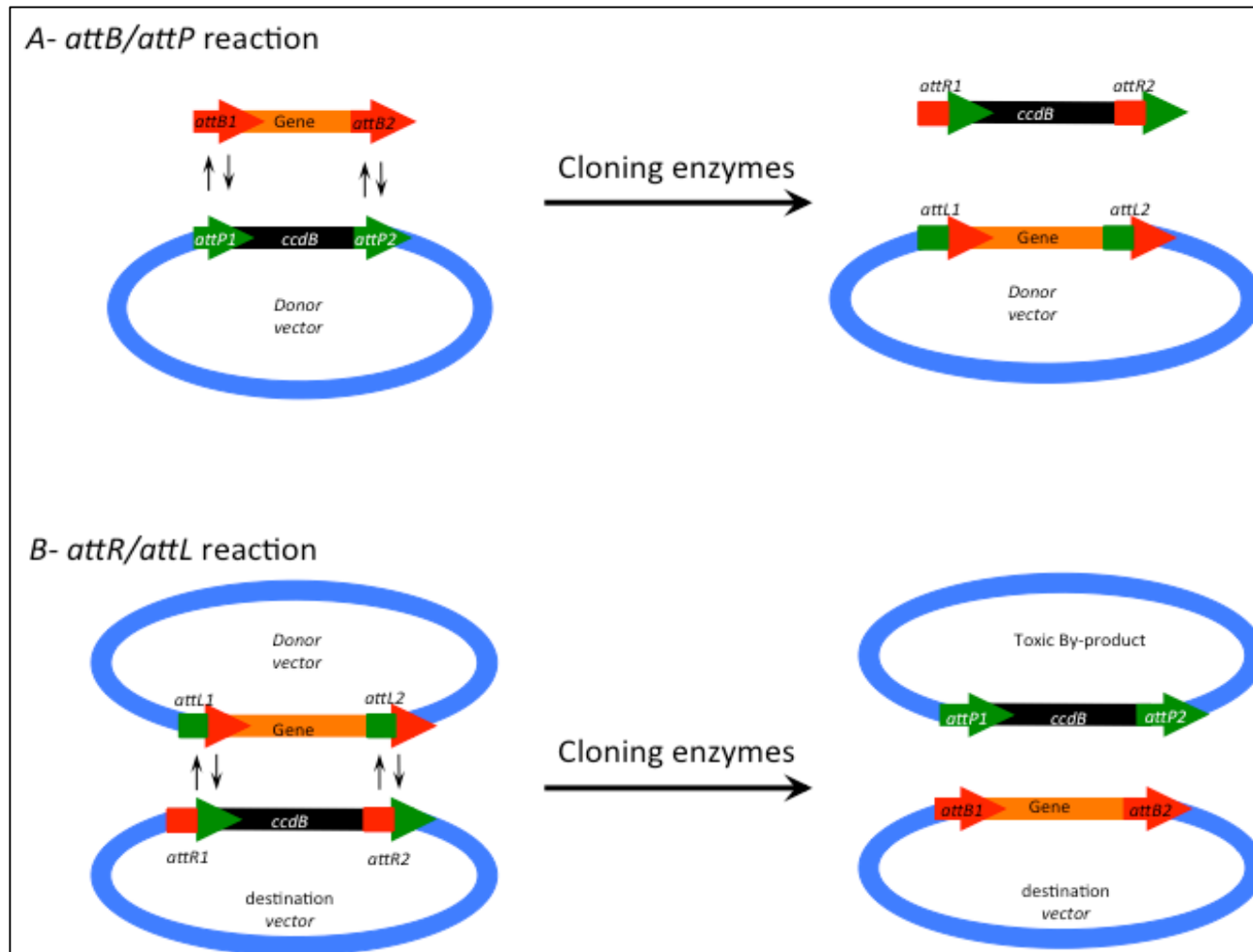


Figure 1.12. The Gateway® cloning system: **A.** the cloning of the interested gene inside the donor or entry vector using the *attB/attP* recombination reaction and that makes the interested gene flanked now with attL1 and attL2 hybrid sites. **B.** The transferring of the interested gene from the donor vector to the destination vector using the attR/attL recombination reaction (Adapted from Liang et al., 2013).

recombination reaction (Fig. 1.12). Gateway® system comes into its own by enabling precise and efficient downstream cloning of the gene of interest into destination plasmids, which are engineered to contain negative selectable marker that are flanked by *attR1* and *attR2*. The negative selectable marker is usually the *ccdB* gene, coding for the toxic protein (CcdB) that acts as a DNA poison, ultimately causing cell death.

The recombination reactions between the donor or entry plasmids and the destination plasmids are performed in a test tube in the presence of a suitable reaction buffer along with purified Int and IHF. The recombinant DNA product is the destination plasmid that now harbors the gene of interest flanked by recombination sites *attB1* and *attB2*. The direction of this approach can be reversed with the presence of the λ Xis protein to restore the original substrates (Esposito et al., 2009).

The utilization of lambda recombinase in the Gateway® system drastically reduces problems associated with traditional cloning *e.g.* time-consuming reactions, reliance on the presence and compatibility of restriction sites, and multiple purification steps. The recombination-based Gateway cloning technique enable efficient parallel transfer of the desired gene from the entry vector into a variety of different living cells protein expression systems, like *E. coli* cells expression system, yeast cellsexpression

system, and mammalian cells expression system (Walhout et al., 2000; Park et al., 2015)

1.4.4.2 Chromosomal integration

Integrase-mediated systems have been utilized to construct chromosomal integration vectors since its first description in the early 1990s for several bacterial species (Atlung *et al.*, 1991). These systems depend on the use of either a single suicide plasmid encoding a selectable marker gene, *attP* and *int* gene, or a two-plasmid system in which an *attP* is carried on a suicide plasmid and transformed to target cells that already harbor a separate integrase expression plasmid. Site-specific recombination can take place with the endogenous *attB* site and can be selected for after transformation (Atlung et al., 1991; Lee et al., 1991a; Lee et al., 1991b). Chromosomal integration is so stable and the reaction can not reverse itself, as the excision requires an RDF and *xis* is not present in the suicide vector used in this approach (Landy and Ross, 1977).

This technique has also been used for the creation of recombinant, attenuated vaccines against *Mycobacteria tuberculosis*, using the strain *Bacillus Calmette-Guerin* (BCG). BCG is an attenuated bovine strain has been successfully used to stimulate immunity against *Mycobacteria tuberculosis*, the primary causative agent of tuberculosis. In order to allow the expressing of foreign antigens in BCG

and increase its immunogenicity, an integrating plasmid harboring the mycobacteriophage L5 *attP* site along with its cognate integrase gene, and expressing a cassette consisting of a promoter region and including the first six codons of the heat shock protein Hsp60 followed by a multiple cloning site. Genes encoding the interesting antigens could then be introduced as fusion protein to be expressed in vivo post-vaccination (Stover et al., 1991).

Another example of the use of integrases to drive chromosomal integration is a plasmid vector that depends on the activity of the recombinase from the phage ϕ CTX. In addition to the recombinase gene, the vector harbors the cognate ϕ CTX *attP* site, a resistance marker gene and a multiple cloning site. This plasmid efficiently directs integration of its recombinant cargo into the chromosome of *Pseudomonas aeruginosa*, when this plasmid is introduced through conjugation (Wang et al., 1995).

Flp recombinases, encoded by *Saccharomyces cerevisiae* (Dymecki, 1996), can facilitate the removal of the entire plasmid backbone, from a plasmid that has its target FRT sites flanking the multiple cloning sites, to create a markerless recombinant strain, which can be an essential feature for some biological or medical applications. A mediated simultaneous or sequential multiple integrations can also be performed using different integrases to integrate different plasmids into their respective *attB* sites. Integrases that are used in this approach are usually

provided from phages that infect bacterial strains intended for genetic manipulation. As most of these assays rely on the presence of targeted *attB* sites for each integrase and the required host-encoded accessory factors (St-Pierre et al., 2013; Rodriguez et al., 2003)

1.4.4.3 Tyrosine integrases in mammalian cells

Lambda integrase has even been successfully used in many applications of genetic engineering in mammalian cells. However, mutations being introduced to integrase gene, makes this integrase so suitable in such applications. Lambda integrase recognizes a big size *attP* site, and it requires a supercoiled DNA and integration host factor protein IHF, to drive DNA recombination. The development of mutants so that the integrase does not need IHF or supercoiling requirements to perform DNA recombination may increase its utility. Two types of mutations were introduced in *int* gene to produce Int-h and Int-h/218. Both of the resulting mutant integrases have been shown to drive its function but without the requirement of accessory factors or supercoiled DNA (Lorbach et al., 2000). The mutant integrase Int-h can recombination in the absence of IHF but at a lower level of function. *In vitro* studies revealed that Int-h prefer supercoiled DNA if IHF present in the reaction buffer, but no preference were showed in the absence of IHF. It was stated that the reason behind the ability of mutant

integrase to work even without the presence of IHF was due to an increased affinity of the enzyme for core-type sites (Patsey and Bruist, 1995).

The Int-h/218 mutant integrase has been used to perform intermolecular recombination in embryonic stem cells chromosomes of mouse. The mutant integrase can, at a low frequency, integrate a cassette that carries a resistant marker in to the genome. The cassette has to have suitable lambda *att* sites for the integration to take place. Excision of the resistance marker, which activated transcription of a green fluorescent protein gene, could be detected by FACS analysis occurred at a frequency of 0.01 – 0.1% (Christ and Dröge, 2002).

The recombinase Flp, from the yeast *Saccharomyces cerevisiae*, and Cre recombinase, from phage P1, are two more examples of the use of integrases in mammalian cells. Both recombinases promote site-specific recombination events between two identical DNA sequences. The target sequence for Flp is the 48-bp FRT site and that for Cre is the 34-bp loxP site. Both Flp and Cre belong to the simple/ bidirectional tyrosine-type recombinase family and do not require host-encoded accessory proteins or specific flanking sequences. This simplicity makes both the Cre/loxP and Flp/FRT systems useful tools for in vivo DNA rearrangements and gene knock-out or knock-in in variety of heterologous environments, including human cells (Hirano et al., 2011; Austin et al., 1981; Broach et al., 1982).

The combination of gene targeting techniques, performed by Flp/FRT and site-specific recombination systems, by Cre/loxP, have been widely exploited to produce gene insertions, deletions, exchanges and inversions in higher eukaryotes cells such as mammalian or mice cells, since both of these integrases can recombine specific DNA sequences with high fidelity and without accessory proteins requirements (Branda and Dymecki, 2004).

1.5. Φ 24B Integrase Protein

The bacteriophage ϕ 24_B is a Shiga toxin-encoding, temperate phage. To date, all characterized Stx phages belong to the lambdoid family. This phage can follow either lytic or lysogenic replication/infection strategies in the *E. coli* bacterial host. When the phage genome is integrated into the host chromosome, it becomes a prophage and gives its host the ability of producing Shiga toxin. This infection converts the *E. coli* into a Shiga-toxin producing *E. coli* (STEC) (Allison, 2007; O'Brien et al., 1984).

The integrase that drives the integration the ϕ 24_B phage genomic DNA into the *E. coli* host chromosome is a phage-encoded protein belonging to the tyrosine family of site-specific integrases. Normally, when lambdoid phages integrate into their host's genome, the prophage confers immunity against other related lambdoid phages being able to infect the lysogen *E. coli* host. However, this immunity model

does not hold for the Stx phage $\phi 24_B$ (Allison et al., 2003) or for other phages that carry an integrase like the one encoded by $\phi 24_B$ (Koudelka et al., 2004). Phage $\phi 24_B$ can integrate into more than five sites within a single bacterial chromosome. This, consequently, increases the pathogenic potential of the infected bacterial host, enabling multiple copies of Shiga toxin genes, which can lead to increased toxin production from the *stx* gene (Fogg et al., 2007; Fogg et al., 2011). This multiple integration takes place inside one bacterial chromosome as $\phi 24_B$ Int has been shown to be able to integrate within four distinct sites within the *E. coli* host. However, these bacterial attachment sites share only about 50% DNA sequence homology. The first recognized site, the primary site, shares 24 bp complete identity with the phage attachment site, *attP* sequence (discussed further in Chapter 5).

The efficiency of new prophage integration increases with each successive superinfection event due to the increased level of Int production. It has been stated that the expression of $\phi 24_B$ Int is enhanced 2.6 fold in the double lysogen and 7.5 fold in the triple lysogen, compared to its expression level in single lysogen (Fogg et al., 2011). Unlike lambda integrase, $\phi 24_B$ Int is not under the control of repressor protein, CI. The orientation $\phi 24_B$ Int gene is inverted compared to integrase genes in other lambdoid phages, which means it is not repressed in the lysogen. In fact, Int has been shown to be constitutively

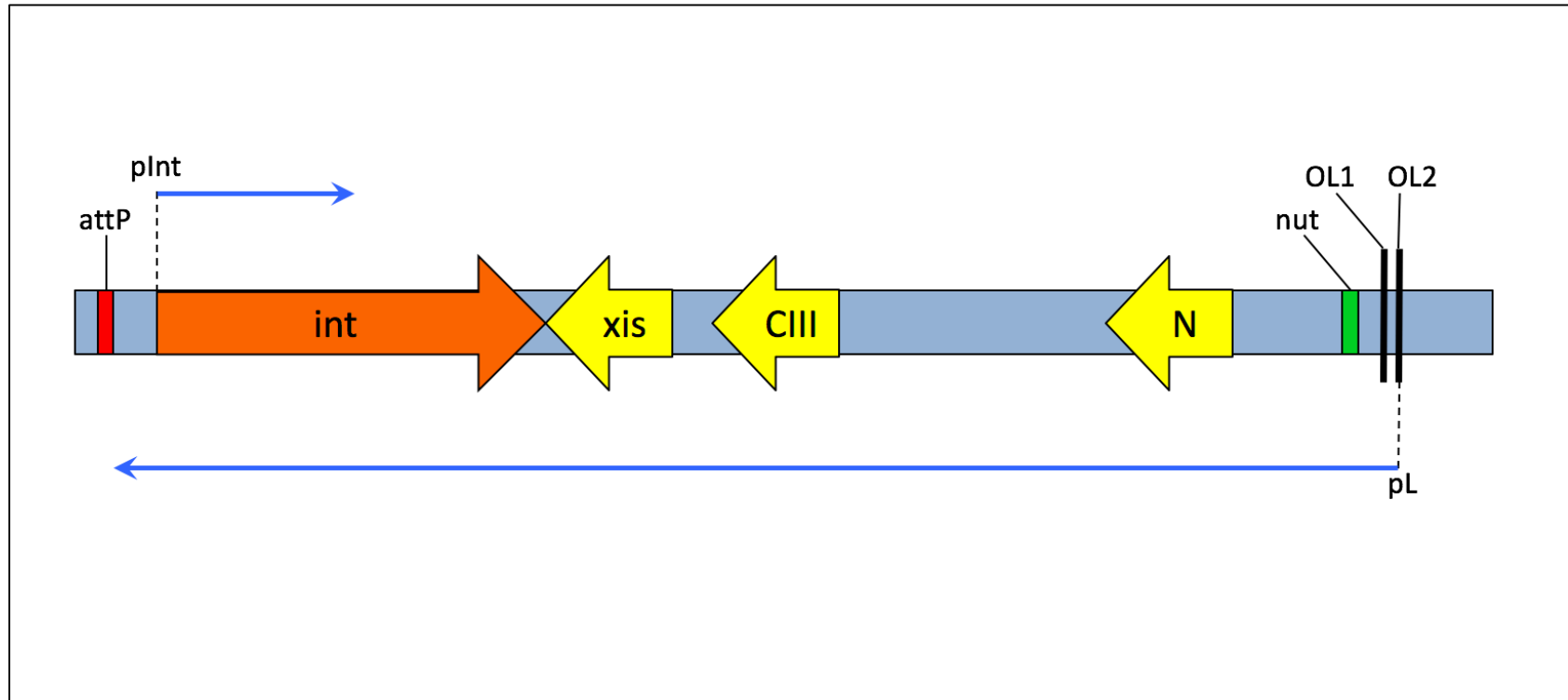


Figure 1.13. The integrase region in the $\phi 24B$ phage genome. This genetic map display the integrase transcription in an inverted orientation compared to its orientation in lambda phage. Because of its orientation, the transcription of $\phi 24B$ Int is not under the control of repressor CI (Adapted from Fogg et al., 2011).

expressed by $\phi 24_B$ lysogens so that active enzyme is always being made (Fogg et al., 2007). Several experiments were performed to confirm this interpretation of the control of $\phi 24_B$ Int expression. Artificial expression of the CI repressor gene by an arabinose inducible promoter and the use of qPCR confirmed that the $\phi 24_B$ Int gene was not repressed though the $\phi 24_B$ Q gene was repressed as expected (Fig.1.13). The Q antiterminator gene expression was used as positive control for this assay (Fogg et al., 2011). However, $\phi 24_B$ forms an inducible prophage, so that the excision of the prophage genome from the *E. coli* chromosome must somehow still be regulated by CI repression. This was indeed found to be the case. Int, itself, though necessary for excision, requires an additional recombination directionality factor. Conforming to the model of tyrosine recombinase activities, the $\phi 24_B$ integrase is responsible for the phage DNA integration inside bacterial chromosome, however, bacteriophage $\phi 24_B$ excision is also catalyzed by Int and an RDF, and this RDF is under control of P_L , which cannot be repressed by CI (Fogg et al., 2011).

1.6.The study aims

The aim of this study is to fully characterize the requirements of the integrase carried by the Stx phage $\Phi 24_B$ (Int $\phi 24_B$) to drive integration, as the four confirmed integration sites in a characterized *E. coli* host share little sequence identity. The strategy for characterizing the requirements for the integrase is to:

1. Establish a working protocol for both an *in vitro* and an *in situ* integration assay based upon the *in vitro* assay used for the bacteriophage Lambda integrase and the working *in situ* Int ϕ 24B assay.
2. Clone the minimal sequence lengths surrounding the characterised primary integration site required for both *attB* and *attP*.
- 3- Determine which nucleotides within the minimal *attB* site are necessary and essential to support integrase recognition through site directed mutagenesis.
- 4-Determine the recombination frequency and preference of each recombination site with the *E. coli* chromosome, as ϕ 24_B integrase can integrate the phage genome inside more than five different bacterial attachment sites *attBs*.
- 5-Determine the importance of bacterial-encoded integration host factor, IHF, in the phage integration.

CHAPTER 2. GENERAL MATERIALS AND METHODS

2.1. Bacterial Strains

All bacterial strains used are listed and described in Table 2.1.1. *E. coli* K-12 strain, MC1061, was used as the source for the *attB₁₋₄* sequences. *E. coli* K-12 strain BL21 was used to produce recombinant IHF, while, *E. coli* K-12 strain TOP10 was used to host all of the *in situ* recombination assays.

Table 2.1.1. Bacterial strains used in this study

| Strain | Description | References |
|----------|--|---|
| MC1061 | <i>E. coli</i> K-12 derivative | {James, 2001, Lytic and lysogenic infection of diverse <i>Escherichia coli</i> and <i>Shigella</i> strains with a verocytotoxigenic bacteriophage; Casadaban, 1980, Analysis of gene control signals by DNA fusion and cloning in <i>Escherichia coli</i> } |
| TOP10 | Competent cells, <i>recA</i> [*] | Invitrogen |
| BL21(AI) | <i>araB</i> ^o ::T7RNAP ^N - <i>tetA</i> ^p , <i>recA</i> ⁻ | Invitrogen |
| JW1702-1 | Δ <i>ihfA786</i> ::kan ³ | Coli Genetic Stock Centre (Baba et al., 2006) |

* *recA*⁻: mutation in DNA repair recombinase gene.

^o*araB*: arabinose inducible promoter.

^NT7RNAP: bacteriophage T7 RNA polymerase.

^p*tetA*: tetracycline resistance gene.

³ Δ *ihfA786*::*kan* : Integration host factor production deficient mutant created by the allelic replacement of the *ihfA* gene with the *aphA* gene conferring resistance to the antibiotic kanamycin.

Table 2.2.1. Plasmids used in this study

| Plasmid | Description | Source |
|------------------------------|---|--------------------------|
| pEE2003 | Derivative of pET16b, Am ^R * | (Frumerie et al., 2005) |
| pBAD/Myc-His C | araBAD, C-myc/His6 epitope, pBR322 ori, Am ^R | Invitrogen |
| pBR322 | pMB1 ori, tet ^R | (Bolivar et al., 1977) |
| pME 4510 | pBR322 ori, Gen ^R | (Rist and Kertesz, 1998) |
| pCR2.1 | TA Cloning, Kan ^R / Am ^R | Invitrogen |
| pΦ24B-int | pBAD carrying 24B int | (Fogg et al., 2011) |
| pACYCDuet TM -1 | P15A ori, Cm ^R | Novagen |
| pCDF Duet TM -1 | CDF ori, Sp/Sm ^R | Novagen |
| pET Duet TM -1 | pBR322 ori, Am ^R | Novagen |
| pRSFDuet TM -1 | RSF ori, Kan ^R | Novagen |
| pCOLADuet TM -1 | ColA ori, Kan ^R | Novagen |
| pET Duet ^{Gen} | pBR322 ori, Gen ^R | This study |
| pCOLADuet ^{tet} | ColA ori, tet ^R | This study |
| pCR-Φ24B-attB ₆₀₀ | pCR2.1 carrying Φ24B attB ₁ (600 bp) | This study |
| pCR-Φ24B-attB ₄₀₀ | pCR2.1 carrying Φ24B attB ₁ (400 bp) | This study |
| pCR-Φ24B-attB ₃₀₀ | pCR2.1 carrying Φ24B attB ₁ (300 bp) | This study |
| pCR-Φ24B-attB ₂₀₀ | pCR2.1 carrying Φ24B attB ₁ (200 bp) | This study |
| pCR-Φ24B-attB ₁₀₀ | pCR2.1 carrying Φ24B attB ₁ (100 bp) | This study |
| pCR-Φ24B-attP ₆₀₀ | pCR2.1 carrying Φ24B attP (600 bp) | This study |
| pCR-Φ24B-attP ₄₀₀ | pCR2.1 carrying Φ24B attP (400 bp) | This study |
| pCR-Φ24B-attP ₃₀₀ | pCR2.1 carrying Φ24B attP (300 bp) | This study |
| pCR-Φ24B-attP ₂₀₀ | pCR2.1 carrying Φ24B attP (200 bp) | This study |
| pCR-Φ24B-attP ₁₀₀ | pCR2.1 carrying Φ24B attP (100 bp) | This study |
| pΦ24B-attB ₁ | pACYCDuet carrying Φ24B attB ₁ (600 bp) | This study |
| pΦ24B-attB ₁₋₃₂₄ | pACYCDuet carrying Φ24B attB ₁ (324 bp) | This study |

| | | |
|-----------------------------|---|------------|
| pΦ24B-attB ₁₋₁₂₅ | pACYCDuet carrying Φ24B attB ₁ (125 bp) | This study |
| pΦ24B-attB ₁₋₁₁₀ | pACYCDuet carrying Φ24B attB ₁ (110 bp) | This study |
| pΦ24B-attB ₁₋₉₃ | pACYCDuet carrying Φ24B attB ₁ (93 bp) | This study |
| pΦ24B-attB ₁₋₆₀ | pACYCDuet carrying Φ24B attB ₁ (60 bp) | This study |
| pΦ24B-attB ₁₋₂₅ | pACYCDuet carrying Φ24B attB ₁ (25 bp) | This study |
| pΦ24B-attB ₂ | pRSFDuet carrying Φ24B attB ₂ (600 bp) | This study |
| pΦ24B-attB ₃ | pETDuet ^{Gen} carrying Φ24B attB ₃ (600 bp) | This study |
| pΦ24B-attB ₄ | pCOLADuet ^{tet} carrying Φ24B attB ₄ (600 bp) | This study |
| pΦ24B-attP ₆₀₀ | pCDFDuet carrying Φ24B attP (600 bp) | This study |
| pΦ24B-attP ₄₂₇ | pCDFDuet carrying Φ24B attP (427 bp) | This study |
| pΦ24B-attP ₃₅₀ | pCDFDuet carrying Φ24B attP (350 bp) | This study |
| pΦ24B-attP ₂₃₇ | pCDFDuet carrying Φ24B attP (237 bp) | This study |
| pΦ24B-attP ₁₄₀ | pCDFDuet carrying Φ24B attP (140 bp) | This study |

*Am^R: Ampicillin resistance gene .

araBAD: arabinose inducible promoter.

His6: C-terminal hexa-histidine tag.

C-myc: C-terminal myc human oncogene peptide epitope.

Kan^R: kanamycin resistance gene .

P15A ori: P15A replicon.

Cm^R : chloramphenicol acetyl transferase gene .

CDF ori: CloDF13-derived CDF replicon .

Sp^R/Sm^R: spectinomycin resistance gene .

pBR332 ori: the pBR322-derived ColE1 replicon.

Gen^R: Gentamicin resistant gene .

RSF ori: the RSF1030-derived RSF replicon;

ColA ori: the COLA replicon from ColA(1) .

Table 2.2.2. Composition of growth media, buffers and antibiotic preparations

| Media/buffers/antibiotics | Composition |
|---------------------------|--|
| LB Agar Plates | 3.7% (w/v) Luria Bertani Agar (Merck KGaA, Darmstadt, Germany) |
| LB Growth Media | 2.5% (w/v) Luria Bertani Broth(Merck KGaA, Darmstadt, Germany) |
| SOC Growth Media | 10% (w/v) Bactotryptone, 0.5%(w/v) Yeast extract, 10mM NaCl, 2.5 mM KCl, 10 mM MgCl ₂ , 10 mM MgSO ₄ , 20 mM Glucose |
| IPTG (100X stock) | 100 mM IPTG (Bioline, ES532-BO25850) |
| Arabinose (100X) | 205(W/v) L-Arabinose (Sigma, # MKBN4207V) |
| X-Gel | 5-broma-4-chliro-3-ndolyl- β -D-galactoside. |
| TfB I Buffer | 30 mM KOAc, 50 mm MnCl ₂ , 100 mM KCl, 10 mM CaCl ₂ , 15% (v/v) Glycerol |
| TfB II Buffer | 10 mM Na MOPS (PH7), 75 mM CaCl ₂ , 100 mM KCl, 15% (v/v) Glycerol |
| TAE (50X stock) | 2 M Tris Base, 1 M Glacial Acetic Acid, 0.05 M EDTA |
| Ampicillin (1000X) | 100 mg ml ⁻¹ Ampicillin sodium salt (Sigma, # BCBG2945V) |
| Kanamycin (1000X) | 50 mg/ml Kanamycin Monosulphate (Melford, K0126) |
| Spectinomycin (1000X) | 100 mg/ml Spectinomycin dihydrochloride (Sigma, 102K0544) |
| Chloramphenicol (1000X) | 12.5mg/ml Chloramphenicol (Sigma, C-0378) |
| Tetracycline (1000X) | 5 mg/ml Tetracycline (Sigma,T8032 |
| Gentamicin (1000X) | 10 mg/ml Gentamicin (Melford,G0124) |
| Lysis Buffer | 50 mM NaH ₂ PO ₄ , 300 mM NaCl, 10mM imidazole, pH adjusted to be 8 using NaOH. |
| Washing Buffer | 50 Mm NaH ₂ PO ₄ , 300 mM NaCl, 20mM imidazole, pH adjusted to be 8 using NaOH. |
| Elution buffer | 50 Mm NaH ₂ PO ₄ , 300 mM NaCl, 250mM imidazole, pH adjusted to be 8 using NaOH. |
| Tricine Sample Buffer 2X | For 50 ml: 5 ml Tris-Cl(1M,PH6.8), 12 ml Glycerol, 4 g SDS, 1.55g DTT, 10 mg Coomassie Blue R250, water up to 50 ml, stored at room temperature. |

| | |
|---------------------------|---|
| 3M Tris-HCl/SDS pH8.4 | 182g Tris base in 300ml dH ₂ O (pH 8.45 by add HCl), water until 500ml, 1.5g SDS , 30ml 5M NaCl. |
| Anode Buffer, 10X | 242g Tris base in 700 ml dH ₂ O (pH 8.45 by add HCl), Q.S. Water 1L. |
| Cathode Buffer, 10X | 179.2 Tricine in 800 ml dH ₂ O (pH 8.3), 50 ml 20% SDS, water until 1L. |
| Coomassi stain solution | 2g Coomassu blue, 450ml Methanol, 450 ml dH ₂ O, 100 ml Glacial acetic acid. |
| Coomassi destain solution | 450ml Methanol, 100 ml Glacial acetic acid, Q.S. 1L dH ₂ O. |

Ampicillin resistance gene (Am^R); araBAD: arabinose inducible promoter; His6:C-terminal hexa-histidine tag; C-myc: C-terminal myc human oncogene peptide epitope; kanamycin resistance gene (Kan^R); P15A ori: P15A replicon; chloramphenicol acetyl transferase gene (Cm^R); CDF ori: CloDF13-derived CDF replicon; spectinomycin resistance gene (SpR/SmR); pBR332 ori: the pBR332-derived ColE1 replicon; Gentamicin resistant gene (GenR); RSF ori: the RSF1030-derived RSF replicon; ColA ori: the ColA replicon from ColA.

2.2. Plasmids, culture conditions, and buffers used in this study

The plasmids used throughout this study are listed in Table 2.2.1, while routinely used growth media, buffers and antibiotics are listed in Table 2.2.2. Both the primary bacterial attachment site and the bacteriophage attachment site were cloned each inside its pCR2.1 plasmid, as we were testing them in an *in vitro* recombination assay. However, in order to drive an *in situ* recombination assay we had to clone phage attachment site and bacterial attachment sites each in its compatible plasmid.

2.3. Bacterial genomic DNA Extraction

An overnight culture of the desired *E. coli* strain or lysogen was processed using the ISOLATE II Genomic DNA Kit (Bioline,London,U.K.) according to manufacturer's instructions. Samples were quantified using NanoDrop (Thermo-Fisher Scientific Inc., MA, USA) spectrophotometry. DNA samples were stored at -20°C.

2.4. Plasmid recovery

2.4.1 Plasmid mini prep

In order to purify plasmid DNA from bacterial cells, the ISOLATE II Plasmid Mini Kit (Bioline, London, U.K.) was used. Five millilitres of an overnight culture of bacterial cells harbouring the plasmid of interest were centrifuged at 11,000 x g for 1 min., and the resulting cell pellets were processed according to the manufacturer's instructions. Cells were resuspended in 250 µL of resuspension buffer P1. The cells were lysed by the addition of 250 µL P2. The sample was inverted 6-8 times for mixing and incubated at room temperature for 5 min. Neutralisation buffer (300 µL) were added and the sample was mixed by inverting 6-8 times. Cell debris was removed from the DNA sample by centrifugation at 11,000 x g for 5-10 min. Miniprep columns (Bioline) were loaded with the sample supernatant and centrifuged at 11,000 x g for 1 min. Columns were washed with washing buffer 1 (PW1) and washing buffer 2 (PW2), before a final centrifugation step, at 11000g for 2 min, to remove any remaining ethanol. The DNA was eluted using an appropriate amount (30-50ul) of pre heated (50 °C) dH₂O.

2.4.2. Plasmid midi prep

QIAGEN midi kit (W. Sussex, U.K.) was used for plasmid midi preparation. Twenty Five millilitres of an over night culture of bacterial cells harbouring the the plasmid of intrest were centrifuged at 6,000 x g for 15 min at 4°C, and samples pellets were processed according to the manufacturer's instructions. Cells were

resuspended in 4 ml of buffer P1. The cells were lysed by the addition of 4 ml of buffer p2, and the samples were vigorously inverted 4-6 times, before they were incubated for 5 min at room temperature. A four millilitres of pre chilled Buffer P3 were added, mixed thoroughly by inverting 4-6 times, and incubated in ice for 15 min. The samples were centrifuged at 20,000 x g for 30 min at 4 °C.

To make them ready to be used, the QIAGEN-tip 100 columns were equilibrated by applying 4 ml of buffer QBT, and columns allowed to empty by gravity flow.

Then, the samples supernatant were applied to the QIAGEN-tip 100, and left to be moved though it by the effect of gravity flow. After that, and by the same way, the QIAGEN-tip 100 were washed with 10 ml of buffer QC for two times before the DNA was eluted using 5ml of buffer QF in to 15 ml vessel. The DNA was precipitated by adding 3.5 ml room-temperature isopropanol and centrifuged at 15000g for 30 min at 4°C. The supernatant was carefully discarded; the pellets were washed with 70 % ethanol and centrifuged at 15000g for 10 min. After the careful decant of the supernatant, the DNA pellets were air-dried for 5-10 min. The DNA redissolved in the suitable volume of appropriate buffer.

2.5. Polymerase chain reaction (PCR)

MyTaq (Bioline, London, U.K.) was routinely used in DNA amplification. When an error free PCR product was needed, the DNA amplification was conducted using the hot start Highly Fidelity Q5 DNA Polymerase (New England Biolabs, Ipswich,

USA). The PCR conditions using MyTaq polymerase started with 2 min initial denaturation step at 95 °C, followed by 30 cycles of denaturation at 95 °C for 10 sec., primer annealing (temperature determined depending on the specific primers Table 2.5.2.) for 10 sec. and ended an extension at 72 °C for 30 sec. per kb of the desired amplified product, and a final extension step at 72 °C for 2 min. On the other hand, the typical thermo cycling PCR program using the Highly Fidelity Q5 DNA Polymerase started with initial denaturation step at 98 °C for 30 sec., followed by 30 cycles of denaturation at the same temperature for 10 sec., primer annealing time was 15 sec at the specific primer's temperatures. The PCR extension took place at 72 °C. Its time was 30 sec. per kb product. All the PCR amplifications were run using Eppendorf (New York, USA) thermo cycler. dNTPs (Bioline, London, U.K.) is one of necessary gradient in the PCR reaction that performed using the Highly Fidelity Q5 DNA Polymerase. Specific primers were purchased from Eurofins Genomics (Wolverhampton, U.K.). DNA templates for 50 μ l PCR reaction were 1-2 μ l of 10-50 ng DNA preparations, or 0.5 μ l -1 μ l of a resuspended bacterial colony in water (100 μ l) (colony PCR). All PCR experiments were performed along with Negative controls, in which water was added instead of DNA templates. PCR products were separated on TAE-agarose gel.

Table 2.5.1. PCR mastermix using Q5 high fidelity DNA polymerase

| Reagent | 5Q polymerase (50 ul reaction) |
|--------------------------|--------------------------------|
| Reaction buffer | 1x |
| dNTPs | 200 μ M |
| Oligonucleotides primers | 0.5 μ M |
| DNA polymerase | 0.02 U μ l ⁻¹ |

Table 2.5.2. Oligonucleotide primers used for PCR amplifications of DNA

| Primers | Sequence (5'-3') | Annealing temp. (°C) | Reference |
|---------------|----------------------|----------------------|------------|
| T7 promoter | TAATACGACTCACTATAGGG | 60 | Novagen |
| T7 terminator | GCTAGTTATTGCTCAGCGG | 60 | Novagen |
| M13 F | GTAAAACGACGGCCAGT | 60 | Invitrogen |
| M13 R | CAGGAAACAGCTATGAC | 60 | Invitrogen |
| pBAD F | ATGCCATAGCATTTTATCC | 60 | Invitrogen |
| pBAD R | GATTTAATCTGTATCAGG | 60 | Invitrogen |

| | | | |
|--------------------------|--|----------------|-------------------------|
| Int F | TTATATCCATTTAACTAAG | 50 | This study |
| Int R | TTAACCGAGCTGTTTA | 50 | This study |
| Ihf A F | GCAGAGCGGCCTTTTTA | 65 | (Frumerie al., 2005) |
| Ihf A R | TCCAGGCATCATTGAGGGA | 65 | (Frumerie al., 2005) |
| ACYCDuetUP1 F | GGATCTCGACGCTCTCCCT | 65 | Novagen |
| pET Upstream | ATGCGTCCGGCGTAGA | 65 | Novagen |
| attP600 F | CTGAAGGATTTGCGTAAACAGC | 60 | This study |
| attP600 R | GATAGTTCCTGCTGGAGTAATCC | 60 | This study |
| attP 427 F | ACAGTCGGCGATGGTGA | 55 | This study |
| attP 427 R | ATCCTGTACACAAGTTTATAA | 55 | This study |
| attP 427 F (PCR cloning) | *CATCACCACAGCCAGGATCC ACAGTCGGCGATGGTGA | 1- 55 2- 60 | This study |
| attP 427 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG ATCCTGTACACAAGTT | 1- 55 2- 60 | This study |
| attP 350 F | CGTGGATCGGTTTAA | 50 | This study |

| | | | |
|--------------------------|--|----------------|------------|
| attP 350 R | GCGTACATATTGAG | 50 | This study |
| attP 350 F (PCR cloning) | *CATCACCACAGCCAGGATCC CGTGGATCGGTTTAA | 1- 50 2- 60 | This study |
| attP 350 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG GCGTACATATTGAG | 1- 50 2- 60 | This study |
| attP 237 F | AGTGGAGTAAAGAACATGC | 60 | This study |
| attP 237 R | CGAACTTACACCTTGATTTTA | 60 | This study |
| attP 237 F (PCR cloning) | *CATCACCACAGCCAGGATCC AGTGGAGTAAAGAACATGC | 1- 60 2- 60 | This study |
| attP 237 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG CGAACTTACACCTTGATTTTA | 1- 60 2- 60 | This study |
| attP 140 F | GGTACTAATATGACGATTAT | 55 | This study |
| attP 140 R | GACGGTATTATCAGTCATA | 55 | This study |
| attP 140 F (PCR cloning) | *CATCACCACAGCCAGGATCC GGTACTAATATGACGATTAT | 1- 55 2- 60 | This study |
| attP 140 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG GACGGTATTATCAGTCATA | 1- 55 2- 60 | This study |

| | | | |
|--------------------------|---|----------------|------------|
| attB 600 F | ATCTGCTTAACGGTGAGCAT | 65 | This study |
| attB 600 R | TTAACGGTACGCTGCACTGG | 65 | This study |
| attB 342 F | GCGTAAAGCAGAAG | 50 | This study |
| attB 324 R | CCATGACGAACTGT | 50 | This study |
| attB 324 F (PCR cloning) | *CATCACCACAGCCAGGATCC GCGTAAAGCAGAAG | 1- 50 2- 60 | This study |
| attB 324 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG CCATGACGAACTGT | 1- 50 2- 60 | This study |
| attB 125 F | CCCCGCACTCCATT | 55 | This study |
| attB 125 R | GCGAACTGATAAATGG | 55 | This study |
| attB 125 F (PCR cloning) | *CATCACCACAGCCAGGATCC CCCCGCACTCCATT | 1- 55 2- 60 | This study |
| attB 125 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG GCGAACTGATAAATGG | 1- 55 2- 60 | This study |
| attB 110 R | GTGTCCCCTGCAGGA | 60 | This study |
| attB 110 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG GTGTCCCCTGCAGGA | 1- 60 2- 60 | This study |

| | | | |
|--|---|--------------------|------------|
| attB 60 F | GGGTATACTCATGCC | 55 | This study |
| attB 60 R | TTAGCCCTTAGGAGG | 55 | This study |
| attB 60 F (PCR cloning) | *CATCACACAGCCAGGATCC GGGTATACTCATGCC | 1- 55 2- 60 | This study |
| attB 60 R (PCR cloning) | *GTTTCTTTACCAGACTCGAG TTAGCCCTTAGGAGG | 1- 55 2- 60 | This study |
| attB 25 F (PCR cloning) | GTCTCTTAGTTAAATGGATATA ACTCGAGTCTGGTAAAGAAAC | 60 | This study |
| attB 25 R (PCR cloning) | TTATATCCATTTAACTAAGAG GAC GGATCCTGGCTGTGGTGATG | 60 | This study |
| Gen ^R gene F | GTCCAACCAGCGGCACCAG | 55 | This study |
| Gen ^R gene R | TTAGGTGGCGGTACTTGGG | 55 | This study |
| Gen ^R gene F (PCR cloning) | *CTAAATACATTCAAATATGTA TCCGGTCCAACCAGCGGCACCAG | 1- 55 2- 60 | This study |
| Gen ^R gene R (PCR cloning) | *GTAAACTTGGTCTGACAGT TACCAATTAGGTGGCGGTACTT GGG | 1- 55 2- 60 | This study |

| | | | |
|--|--|-----------------------------|------------------|
| Tet ^R gene F | TTCTCATGTTTGACAGC | 55 | This study |
| Tet ^R gene R | TCAGGTCGAGGTGGCC | 55 | This study |
| Tet ^R gene F (PCR cloning) | *GGGAAATGTGCGCGGAA TTCTCATGTTTGACAGC | 1- 55 2- 60 | This study |
| Tet ^R gene R (PCR cloning) | *CGCTCATGAGACAATAACCC TGTCA GGTTCGAGGTGGCC | 1- 55 2- 60 | This study |
| attPMut. (2.17. Site-directed mutagenesis) | GTATTTAAATCAATGAGTTAAT ACTTGTATTCATTATATCCATTT AACTAAGAGGACGTTGTTGC, | 60 | This study |
| attPMut. R (2.17. Site-directed mutagenesis) | GGATATAATGAATACAAGTATT AACTCATTGATTTAAATA CATAAAGCACCGAAATAATCGTC | 60 | This study |
| qPCR primers | | | |
| Primers | Sequence (5'-3') | Annealing temp. (°C) | Reference |
| Spect F | TCGCCTTTCACGTAGTGGAC | 53 | This study |

| | | | |
|---------|-------------------------|----|------------|
| Spect R | CGCAGTAACCGGCAAAATCG | 53 | This study |
| RL1 F | CGCACTCCATTAGCGGGTAT | 53 | This study |
| RLR | CGTGGATCGGTTTAA | 53 | This study |
| RL2 F | TGCATCCGGTTACTTGTTGGT | 53 | This study |
| RL3 F | GTGATGACTCGTGCTTAGATCAA | 53 | This study |
| RL3 R | ACAGTCGGCGATGGTGA | 53 | This study |
| RL4 F | CGGGTGGCGATTTCCACGG | 53 | This study |
| RL4 R | GACGGTATTATCAGTCATA | 53 | This study |

These primers were used first to amplify the insert sequence using annealing Temp. 1. The

annealing Temp.2 was used in the amplification of the plasmid backbone in the PCR cloning.

2.6. Agarose gel electrophoreses

TAE-agarose gel electrophoresis was used to separate DNA. The gel concentrations were 0.7%, 1% or 1.5 % depended on the DNA size, and they were run at 100 V and 400 mA, unless otherwise stated. The DNA sample were run along with the Hyperladder I (Bioline, London, U.K.) molecular weight marker. The gels were viewed under UV trans-illumination at a wavelength of 302 nm.

2.7. Recovery of the DNA from the agarose gel

The desired DNA band was cut out from the gel after they were viewed using the blue light box. The gel slice was processed using PCR and gel extraction kit from (Bioline, London, U.K.). DNA was eluted in 25 μ l of water.

2.8. Restriction endonuclease digestion

All restriction digestions were done in final volume of 20 μ l composed of 1 μ l of the appropriate enzyme, 2x of its recommended proprietary buffer, and 1 μ g of the DNA. The restriction enzymes that were used in this study were STAR activity enzymes (Fermantus, MA, USA), incubated at their recommended temperature for 1h. The resultant restriction reaction was separated on a TAE agarose gel and the desired DNA band was recovered.

2.9. Ligation of DNA fragments

The final volume of ligation reaction was 20 μ l composed of 1:1 or 1:3 of plasmid and insert molar ratios. Plasmids and inserts were incubated with T4 DNA ligase (T4 DNA Ligase, Fermantus, #EL0011) and its ligase buffer for 10 min at 22 °C.

2.10. Production of electrocompetent cells

LB broth (400 mL) was inoculated with 1 ml of an overnight culture. This was incubated at 37 °C with shaking at 200 r.p.m to OD₆₀₀= 0.5-0.8. The culture was

placed in ice and cells harvested by centrifugation at 5,000 x g for 10 min. The pellet was resuspended carefully in 1/10 vol of ice-cold ddH₂O. After 3 repeats of the wash step, the cells were finally resuspended in 10% ice-cold glycerol. Aliquots of 100 μ l were flash-frozen in liquid nitrogen and stored at -80°C (Sambrook et al., 1989)

2.11. Transformation of electrocompetent cells

Plasmid DNA (200-400 ng) was added to 100 μ l chilled competent cells. The mixture of cells and plasmid DNA was transformed to a disposable cold 1 mm aperture electro cuvette. Electroporation was performed at 2.50 mV, 200 Ω and 25 μ F. Cells were diluted in 300 μ l of SOC broth media and incubated at 37°C for one hour. Aliquots of 100 μ l were flash-frozen in liquid nitrogen and stored at -80°C (Dower et al., 1988; Sambrook et al., 1989).

2.12. Preparation of chemically competent cells

LB broth 400 ml was inoculated with 1 ml of an overnight culture. This was incubated at 37 °C with shaking at 200 r.p.m to the OD₆₀₀ 0.5-0.8. The culture was placed in ice and cells harvested by centrifugation at 5000 g for 15 minutes. The pellet was resuspend carefully in 100 ml of ice cold TfBI Buffer (Table 2.2.2) and incubated on ice for 30 minutes. Cells were collected by centrifugation at 5000 g for 10 minutes and resuspended in 20 ml of TfBII (Table 2.2.2) and incubated for

a further 30 minutes on ice. Aliquots of 100 μ l were flash-frozen in liquid nitrogen and stored at -80°C (Inoue et al., 1990; Sambrook et al., 1989).

2.13. Transformation of chemically competent cells

A plasmid DNA 200-400 ng was added to 100 μ l of ice thawed competent cells incubated on ice for 30 minutes. The cells were subsequently heat-shocked at 42°C for 45 seconds, immediately diluted 1:10 in SOC broth media and recovered at 37°C for one hour. After incubation, 150 μ l and 50 μ l were spread plated onto LB agar plates with plasmid specific antibiotic. The plates were incubated at 37°C for overnight (Inoue et al., 1990; Sambrook et al., 1989).

2.14. Colony PCR

Colony PCR is a convenient high-throughput method for determining the presence or absence of insert DNA in plasmid constructs. Individual transformants can be lysed during the initial heating step. This initial heating step causes the release of the plasmid DNA from the cell, so it can serve as template for the amplification reaction. Primers designed to specifically target the insert DNA can be used to determine if the construct contains the DNA fragment of interest.

2.15. Protein quantification: Bicinchoninic Acid assay (BCA)

The assay (Smith et al., 1985, Wiechelman et al., 1988) was performed using BAC kit from thermo- fisher scientific (MA, USA). Standards curve of bovine serum albumin ($5 \mu\text{g ml}^{-1}$ – $2000 \mu\text{g ml}^{-1}$) were prepared in triplicate. The BCA working reagent was made by mixing 50 volumes of reagent A [BCA- Na_2 (1%), $\text{Na}_2\text{C}_4\text{H}_4\text{O}_6$ (0.16%), $\text{Na}_2\text{CO}_3 \cdot \text{H}_2\text{O}$ (2%), NaOH (0.4%), NaHCO_3 (0.95%)] with 1 volume of reagent B [$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (4%)]. Three types of dilutions were performed for the samples (1 in 10, 1 in 50 and 1 in 100) in PBS using 96 well plates. Twenty five microliters of each standard or sample (both neat and the three dilutions) were added to individual wells of the microtiter plate. The assay was prepared by adding $200 \mu\text{l}$ of the working reagent to each tube of diluted or undiluted protein solutions and mixed by shaking. The standards and samples plates were incubated at 37°C for 30 min, before the absorbance was measured at 562 nm. Depends on the calibration curve that plotted using the average measurements of standards; the samples quantity of protein was determined.

2.16. The Sequencing analysis

GenBank database searches were employed using blastn, blastp and tblastn (<http://www.ncbi.nlm.nih.gov>). Sequence manipulation, primer binding sites and restriction site analyses were performed using the program Serial Cloner 2.6-1 (http://serialbasics.free.fr/Serial_Cloner.html). Each plasmid construct and all DNA recombination products were sequenced by GATC-Biotech (Konstanz, Germany).

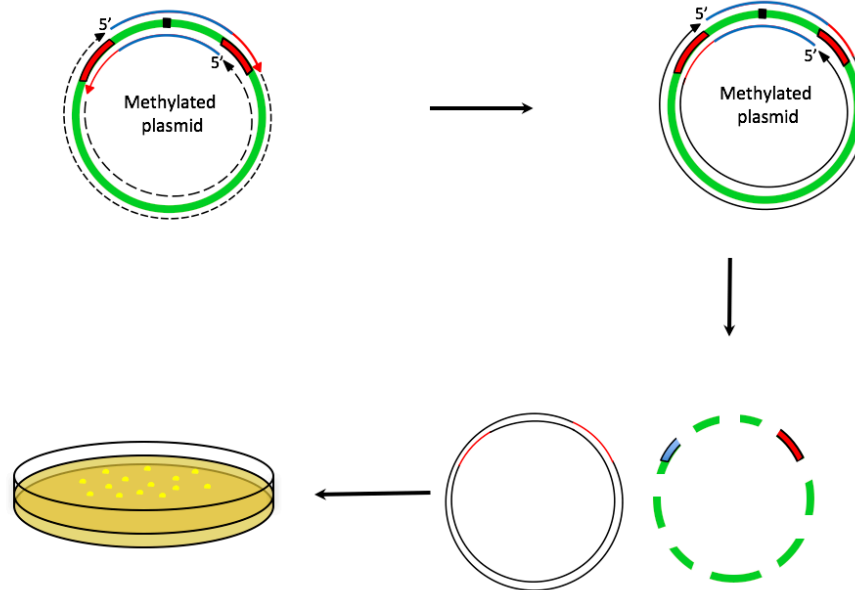
2.17. Site-directed mutagenesis

Site-directed mutagenesis technique (Ke and Madison, 1997) was used to repair any undesired mutation might occur during PCR and plasmid cloning. A DNA mutation of two base pairs in the left flanking DNA sequence of $\Phi 24_B$ *attP* site was repaired using this technique. The long oligonucleotide primers; Forward, GTATTTAAATCAATGAGTTAATACTTGTATTCATTATATCCATTTAACTAAGAGGACG TTGTTGC, and Revers, GGATATAATGAATACAAGTATTAACCTCATTGATTTAAATA CATAAAGCACCGAAATAATCGTC (Wolverhampton, U.K.) (Table 2.5.2), possessing a 40 bp region of complementary overlap were used to repair that mutation in a one-step PCR reaction using the hot start Highly Fidelity Q5 DNA Polymerase (New England Biolabs, Ipswich, USA). The mutant base pairs that needed to be repaired were located in the centre of the primer overlap area (Fig. 2.1). The PCR reaction template was the plasmid clone that had the 2 bp mutation. The typical thermo cycling PCR program started with initial denaturation step at 98 °C for 30 sec., followed by 30 cycles of denaturation at the same temperature for 10 sec., the primer annealing time was 15 sec. at 60 °C. The PCR amplifications were run using an Eppendorf (New York, USA) thermo cycler Model.

After the PCR reaction was completed, the PCR product was purified using the PCR and Gel extraction kit from (Bioline, London, U.K.). The DNA was eluted with 10 μ l of pre-heated water (50°C) before the spin column was centrifuged at 11,000 g for 1 min, as the final step of that purification (See 2.6.1 Recovery of the DNA from the agarose gel, Chapter 2).

1-The long oligonucleotide primers with two domains. They shared an overlap area (Highlighted with blue color). The second domain (red color) in each primer has complementarity to the plasmid (Highlighted with the same colors in the plasmid). The primers were mixed with the methylated plasmid (Green colour), purified directly from an *E. coli* clone culture, that has the DNA mutation, in one step PCR reaction. The mutated nucleotides had to be in the centre of the primer overlap area (Highlighted with black color).

2- PCR extension from the 3' end of the primers were performed by Q5 DNA polymerase until its reach the 5' end of the primer extending around the entire plasmid, in a linear PCR amplification.



4- The new plasmid was then transformed into *E. coli* cells

3-The final PCR product and the original plasmid molecules were digested with *DpnI* digestion enzyme. *DpnI* destroys the methylated template DNA only.

Figure 2.1. Site-directed mutagenesis technique (Adapted from Ke, 1997). In order to repair any mutation within a plasmid clone, long PCR primers, sharing an overlap region of 40 bp, were used in a one-step PCR reaction. The mutated nucleotides had to be located in the centre of the primer overlap area. The PCR product was then treated with *DpnI* digestion enzyme to destroy the methylated template DNA that possesses the mutation only, and then the newly repaired plasmid was transformed into *E. coli* cells.

In order to digest the methylated template plasmid, 8.5 µl of the restriction free cloning reaction were treated with 0.5 µl of *Dpn* I (Bioline), and 1 µl of CutSmart Buffer (Bioline, # B7204S) for 2 h at 37 °C. Then, 5 µl from the reaction were transformed into new Top 10 Competent cells and transformants were selected for on LB agar with the plasmid resistance antibiotic at its specific concentration (Invitrogen). Positive cloning was confirmed by plasmid prep and DNA sequencing (Fig. 2.1).

2.18. Statistical analysis

The *t*-test analysis was used to find out the significance of the qPCR results, Chapter 5.

CHAPTER 3. IN VITRO RECOMBINATION

3.1. Background

3.1.1. Site-Specific Recombinases

At the point of lambdoid bacteriophage infection, when the phage DNA enters the bacterial cell, a number of specific factors must be expressed in order to direct the incoming phage into either the lytic or lysogenic replication cycle (Hendrix, 1983). For the phages destined to become integrated, non-self-replicating prophages, the incoming phage must produce an enzyme that directs recombination with the bacterial chromosome. For lambdoid phages this recombinase is known as integrase.

3.1.2. Lambda phage Integrase

Lambdoid phages typically encode an integrase (a tyrosine, site-specific recombinase (see also 1.4 Site-specific recombination, Chapter 1)) enzyme to drive recombination with the bacterial chromosome and a repressor protein to prevent the expression of the lytic cycle-associated genes (Dodd, 2001). For the integrase to direct prophage integration, two DNA molecules (the circularised phage genome and the bacterial chromosome) are fused together by the integrase enzyme (Fig. 3.1). The phage DNA attachment site is within a region called *attP*, while the complementary site in the bacterial genome is called *attB* (Mizuuchi and Mizuuchi, 1980).

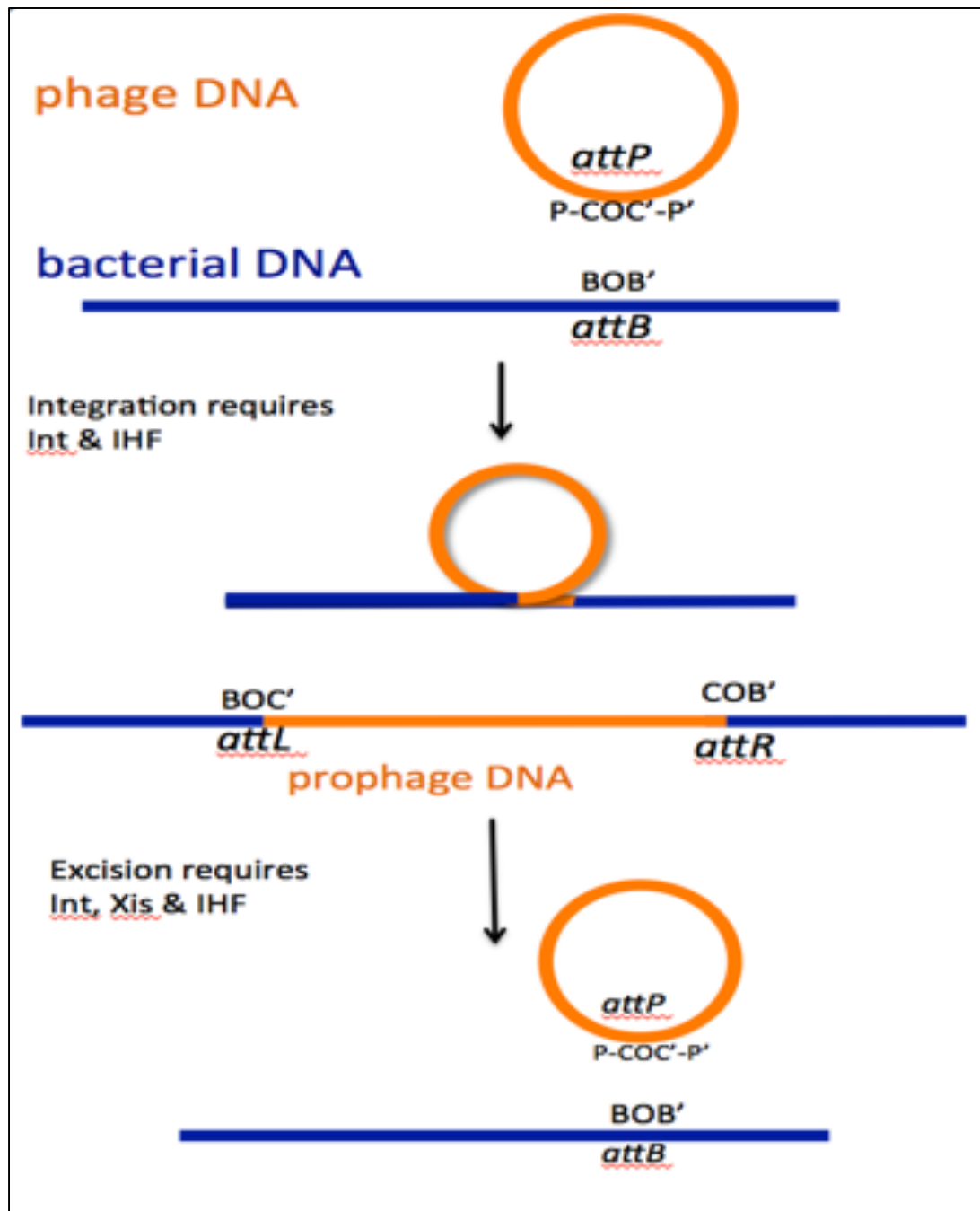


Figure 3.1. Circular lambda phage DNA is converted to an integrated prophage by a reciprocal recombination between *attP* and *attB*. The attachment site of the phage chromosome (*attP*) is composed of two integrase binding sites (P,P') and the core binding site(COC'); while the attachment site of the bacterial chromosome (*attB*) is composed of two core binding sites (B,B') and the overlap region (O); Int, phage encoded integrase; Xis, phage encoded excisionase also referred to as a recombination directionality factor; IHF, host cell-encoded integration host factor.

DNA recombination starts with formation of a DNA Holliday junction, which is then resolved at that junction to create new two sites, *attL* and *attR* (Stark et al., 1992). This process needs a short *attB* overlap region, or O, flanked by short imperfect inverted repeats B and B', and an extended *attP* sequence with multiple integrase binding sites as well as binding sites for accessory cofactors (Fig. 3.1). For the archetypal lambdoid phage, Lambda, it has been discovered that *attP* is composed of two of integrase arm binding sites (P and P') and the core-binding site (COC'). Furthermore, for lambda phage integration, IHF binding sites exist between COC' and the P. On the other hand, for lambda phage, the bacterial chromosome *attB* site is composed of core binding sites B and B' flanking O region (Frumerie et al., 2008) (Fig. 3.1).

When the *attB* and *attP* sites are cut by the integrase, between the B and O regions in the bacterial genome and the P and O region in the phage chromosome, respectively, the bacterial and phage genomes are recombined to produce new *attL* and *attR* sites. These sites can later be utilized by the integrase and its cognate recombination directionality factor for excision of the phage genome from the bacterial chromosome (Sarkar et al., 2009). Early studies of site-specific recombination systems showed that DNA homology in the DNA crossover region is necessary to promote recombination, and it identified that the crossover regions interact with each other through complementarity via DNA-DNA interactions in addition to DNA-protein interactions enabling site-specific recombination (Weisberg et al., 1983).

Recombination catalyzed by lambda integrase and most tyrosine recombinases takes place by sequential exchanges of the crossover regions of the *attB* and *attP* sites. Four recombinase monomers bind to the DNA substrate at the same time (Echols and Green, 1979), when one strand in each site is cut by the first two active monomers to produce a 3'-phosphotyrosyl intermediate and a free 5'-hydroxyl group. The next step involves a DNA strand exchange and rejoining reaction when the cleaved strand in each site migrate to allow the 5'-hydroxyl groups to attach the phosphotyrosyl bond of the opposite site to form a phosphodiester bond reforming the integrity of the DNA strand. This generates a Holliday junction (HJ) intermediate (Biswas et al., 2005). The HJ is resolved when the second pair of the integrase monomers activated to carry out the second cut of DNA followed by the second exchange of the DNA strand and ligations results in the recombinant (Hsu and Landy, 1984). This recombination model is called the "branch migration" model (Weisberg et al., 1983)(Fig. 1.10, Chapter 1). Homology between two sites is important for branch migrations (Echols and Green, 1979).

Upon conditions that induce the host cell's SOS repair system, for example irradiation with ultraviolet light, or DNA damage, the prophage is induced resulting in excision of the integrated prophage from the lysogen's chromosome and entry of the excised phage genome into the lytic replication cycle. This results

in the production of new bacteriophages that are released after the burst of bacterial host cell (Hendrix, 1983)

Int plays a central role in prophage excision, though; a phage- encoded recombination directional factor RDF is additionally required to control the reaction directionality. In the lambda excision system, both the phage-encoded excisionase (Xis) and Integration host Factor (IHF) are the RDF proteins that Int needs to drive excision. In addition, lambda phage excision is usually enhanced by the factor for inversion stimulation (Fis)(Numrych et al., 1992).

As we have mentioned above, there are two types of lambda Int binding sequence, core type and arm type-sites. There are five arm-binding sites located outside the crossover region on *attP*, *attR* and *attL*. Three neighboring sites are located on *attP* and *attL*, called P'1, P'2 and P'3. The two other sites are called P1 and P2 and located on *attP* and *attR* (Fig.3.2)(Cho et al., 2002).

Footprint experiments showed that each of these RDF proteins, Int and IHF binds to a specific sequence within *att* sites (Fig 3.2). Int binds to all of the five arm binding sites and to two more binding regions in each core (B and B' in *attB* and, C and C' in *attP*). IHF binds to one site in P' and in to sites in P within *attP*. And within *attP* also, Fis binds to F site, which overlaps X2 (one of two Xis sites) (Numrych et al., 1992).

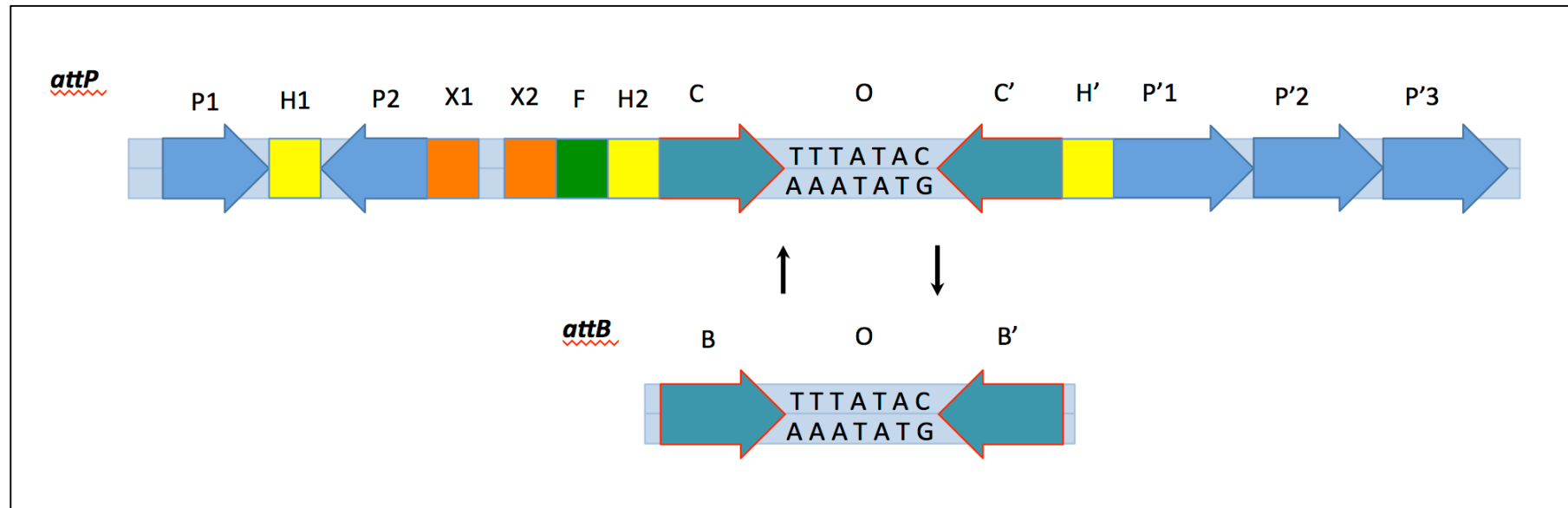


Figure 3.2. Details of the Lambda *attP* and *attB* sites(Numrych et al., 1992). The overlap region between them is 7 bp, flanked by inverted repeats of 7 bp (C,C',B, and B'). The Int arm type binding sites are P1, P2, P'1, P'2 and P'3. There are three IHF binding sites (H1, H2, and H'), one Fis binding site (F) and two Xis binding site.

Lambda Xis, a recombination directionality factor, identifies two direct, imperfect, 13-bp repeats, called X1 and X2, which exist between the P2 site and the crossover region on *attR*. Xis binding to X1 and X2 facilitates a bend in the DNA strands and promotes Int-DNA binding. The binding between Int and P2 site is so weak that it requires Xis to create the *attR* nucleoprotein complex (Thompson et al., 1987; Yin et al., 1985). It has been reported that Xis the carboxyl-terminal region has a direct contact with Int protein, while the C-terminal region has 53 amino acids that specifically bind to the Xis sites in the DNA, and have no ability to interact with Int (Numrych et al., 1992).

3.1.3. Integrase of $\Phi 24_B$

Bacteriophage $\Phi 24_B$ is a member of the lambdoid family (Fattah et al., 2000). The integrase encoded by $\Phi 24_B$ phage is a ~45 kDa protein, encoded by the *int* gene (EF397940.1). $\Phi 24_B$ integrase protein belongs to the tyrosine recombinases family (Balding et al, 2005). As we already have seen in the introduction, all members of this family use a tyrosine amino acid to derive DNA cleavage and integration.

Int possesses high affinity binding for the *attP* site-arm regions and low affinity to the *attP* core-binding region (Richet et al., 1988). Its recombination mechanisms take the form of a tetrad complex composed of four Int monomers; each monomer

has an N-terminal domain that resides over each adjacent core-binding region (Biswas et al., 2005). The C-terminal domains (core binding and catalytic domains) recognize and bind to the identical crossover sequences in the *attP/attB* regions in a square planar formation, promoting DNA recombination (Van Duyne, 2005). Lambda integrase is the only phage-encoded protein needed for lambdoid phage genome integration within the host genome (Hendrix, 1983).

However, most tyrosine integrases do not act autonomously, they also require the help of bacterial encoded accessory factors (Richet et al., 1988). The accessory factor for Lambda integrase is known as integration host factor IHF. IHF interacts with the DNA sequences surrounding the attachment site in order to expose the DNA to the action of integrase that cuts the DNA using its topoisomerase-like activity (Sugimura and Crothers, 2006).

IHF is a ~ 21.8 kDa DNA-binding protein composed of two monomers (Sanyal et al., 2014; Azam and Ishihama, 1999). It is a small heterodimeric protein composed of two subunits, α and β , which are encoded by bacterial genes (Yang and Nash, 1995). The subunit α is encoded by the *himA* gene and composed of 300 bp. The molecular weight of the α subunit is ~11,000 kDa. The gene that is responsible for the β subunit is *himD*; this gene is 280 bp in length, and encodes a protein product of ~9,500 kDa. The active component of IHF is the heterodimer of both α and β subunits (Kikuchi and Nash, 1978).

IHF protein works as an accessory protein that enhances the capacity of Int's enzymatic activity. IHF is a site-specific DNA binding protein, and there are three loci surrounding the lambda *attP* that support a high affinity interaction with IHF. There are, however, specific IHF binding sites located within non-*att* DNA sequences in the *E. coli* genome. Most of these sites are adjacent to genes whose expression is altered in IHF mutants (Craig and Nash, 1984). The IHF-DNA complex is unique among binding proteins. This protein typically binds to a 30 - 35 bp DNA sequence that is composed of at least two main binding domains. Alignment among IHF binding sequences within an already created *attL* population library showed that the 3' region is where IHF binds strongly and specifically. This region can be identified as a 13 bp sequence of WATCAANNNTTR. Unlike the 3' region, the 5' region is random in most cases. However the 5' sequences are all typically AT rich, but without any obvious sequence patterns (Goodman et al., 1999).

In addition to its role in site-specific, phage-genome integration into the *E. coli* genome, IHF is a necessary protein for a wide range of cellular activities including: site-specific DNA recombination, DNA replication, phage packaging, plasmid partitioning, and the expression of some genes like: type 1 fimbriae of *E. coli*, encoded by *fimA* (Friedman, 1988; Arfin et al., 2000; Le et al., 2013).

Lambda integrase can recognize complex binding sites in both the phage and host bacterial genomes (Richet et al., 1988). The λ phage attachment site, *attP*, is a quite complex site composed of several adjacent binding sites called arm-type sites that flank the core-type sites (Ross et al., 1979). Furthermore, it contains other binding sites for factors involved in DNA integration and excision (Yin et al., 1985). On the other hand, bacterial attachment site for λ phage, *attB*, is less complicated, composed of two core-type binding sites, which have an inverted repeat sequences flanking the overlap region. Crossing over usually occurs between the completely identical overlap regions in both *attB* and *attP* {Sarkar, 2001, The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions}.

3.1.4 Aims

- To over express and purify both INT and IHF proteins.
 - Integrase enzyme, INT, is necessary protein to run DNA recombination. However, it does not normally function alone.
 - IHF is a required cofactor for the function of most characterized tyrosine recombinases.
- To clone of both *attB* and *attP* sequences in separate plasmids.
- To establish a working protocol for an *in vitro* integration assay based upon the *in vitro* assay used to characterize the bacteriophage Lambda.

3.2. Specific Methods

3.2.1. Cloning of bacterial and phage attachment sites

The *attB* core and either 100 bp, 300 bp or 600 bp of flanking sequence were amplified using *attB*100 F/R, *attB*300 F/R or *attB* 600F/R primers, respectively (See 2.5 PCR, Chapter 2) (Table 2.2.1). Each of these *attB* sequences was cloned, separately, into the pCR2.1 cloning vector (Invitrogen), according to the manufacturer's guidelines. At the same time, the *attP* core and either 100 bp, 300 bp or 600 bp of flanking sequence were amplified using *attP*100 F/R , *attP*300 F/R or *attP* 600F/R primers, respectively (See 2.5 PCR, Chapter 2) (Table 2.2.1). Each of these *attP* sequences was cloned, separately, into the pCR2.1 cloning vector (Invitrogen), according to manufacturer's guidelines. The three *attB* plasmids, along with three *attP*-cloned pCR2.1 plasmids, were all individually transformed into Top 10 chemically competent cells. The transformants were cultured overnight on LB agar plates containing ampicillin (100 $\mu\text{g ml}^{-1}$) and kanamycin (50 $\mu\text{g ml}^{-1}$). The identity of the desired transformants were confirmed by colony PCR (See 2.13 colony PCR, Chapter 2). Freezer stocks were made from each positive transformant and stored at -80 °C (See 2.12. Transformation of chemically competent cells, Chapter 2).

3.2.2. Protein Overexpression

Overnight LB broth cultures (1 mL) of cells harboring the interested plasmid were used to inoculate 10 mL of LB broth. This subculture was incubated at 37 °C with

shaking at 300 r.p.m to the $OD_{600} = 0.5-0.7$. A sample (1 mL) was immediately taken from the culture; the cells were recovered by centrifugation for 20-30 min, at 10,000 g at 4 °C, and resuspended in 50 μ l of Tricine Sample Buffer (Table 2.3), and was frozen at -20 °C for later use as the non-induced control cells. The rest of the culture was induced with the appropriate plasmid inducer to the appropriate final inducer concentration (0.02 % w/v final concentration of L-Arabinose was used to induce the overexpression of Int from p Φ 24B-int, while the final concentration of IPTG for overexpression of IHF was 5 mM (Table 2.2.1)). The induced culture was incubated at 37 °C with shaking at 300 r.p.m. for a further 3-4 hrs. The culture was then placed on ice and a sample (1ml) was taken, the cells in that 1 mL harvested by centrifugation for 20-30 min, at 10,000 g at 4 °C, and then resuspended in 100 μ l of Tricine Sample Buffer (Table 2.3), and finally frozen at -20 °C until needed (induced control). The rest of cells in this culture were then harvested by centrifugation for 20-30 min at 10,000 g at 4 °C. The resultant pellet was resuspended carefully in 5 ml of ice-cold lysis buffer for purification of the His-tagged protein under native conditions. Using a sonicator with a microtip probe, the sample was sonicated 6-8 times with 10 seconds pauses at 200-300 Watt. The lysate was kept on ice at all times, until it was subjected to centrifugation for 20-30 min at 10,000g at 4°C. The supernatant was harvested (crude cell extract A, soluble proteins) and stored on ice. The resultant pellet was resuspended in 5 ml lysis buffer (crude extract B, insoluble proteins).

Crude extracts A and B (5 μ l, each), were mixed with 5 μ l of Tricine Sample Buffer (Table 2.3). These samples, along with the noninduced and induced control samples, were heated at 95 °C for 5 min. The samples were then centrifuged at 15,000g for 1min, before being loaded (20 μ l from each) on a 12% SDS-PAGE gel. The gel was run at 150 V min⁻¹ for 30 min.

3.2.3. SDS PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)(Laemmli, 1970), was used to examine the protein preparations, using the Mini-Protean 3 Unit (Bio Rad, Serial NO 67S). The size of the vertical running gel was 75 mm and composed of a bottom Separating gel (1 ml water, 1.25 ml 3M Tris-HCl/SDS pH8.4, 1.12 ml 40% acrylamide, 0.37 ml glycerol, 5 μ l 30% APS, and 5 μ l TEMED), and a Stacking gel (2.03 ml water, 0.77 ml 3M Tris-HCl/SDS pH8.4, 0.31 ml 40% acrylamide, 5 μ l 30% APS, and 5 μ l TEMED). The separating gel mixture was poured first in between the 75 mm glass plates (Bio-Rad, #1653308), and covered with isopropanol until the gel polymerised (20-30min). Then, the isopropanol was washed with water, dried, and the stacking gel mix was poured over the top of Separating gel and the 20 μ l-size comb was used to cover the top of the stacking gel until is polymerised. The fully polymerized gel was placed into the running unit, and the middle tank was filled with an appropriate volume of cathode buffer whilst the outside area of the unit was filled with anode buffer (Table 2.2.2). The comb was removed and 20 μ l (15 μ l sample and 5 μ l Tricine sample buffer) of the bold samples were loaded in the gel in individual wells along with a prestained molecular weight marker. SDS-PAGE was performed a 150 V min⁻¹ for 30 min.

The unit was then dismantled and the gel was placed in Coomassi Blue Stain solution (Table 2.3) with gentle agitation for 30 min. The gel was destained by replacing the stain solution with Coomassi Blue Destain solution (Table 2.3). The destain solution was changed several times until a clear gel background was obtained.

3.2.4. Semi-Dry Western blot

All western blot buffers are listed in Table (3.2.1). Whilst the SDS-PAGE gel was running a piece of nitrocellulose transfer membrane (Thermo Fisher Scientific) and extra thick blot paper (Bio-Rad, 1703969) were cut to the size of approximately 6 cm × 8 cm. The extra thick blot paper was pre wet for 5-10 minute in blotting buffer (Table 3.2.1).

When the unstained gel become ready after SDS gel running was done; the blotting sandwich was assembled on the blotter as follow:

- The extra thick blot paper presoaked in blotting buffer.
- The SDS-Page gel.
- The nitrocellulose transfer membrane
- Another piece of the extra thick blot paper presoaked in blotting buffer.

The blot was run using the Trans-Blot SD semidry Transfer cell (Bio-Rad), for 1h at 78 mA and 130 V. Afterwards, the blot was blocked in blocking buffer (Table 3.2.1) for at least 1 hour at 25° C, or overnight at 4° C, as appropriate, with gentle agitation. The blot was washed twice with TBST buffer for 5 min before being incubated in primary antibody solution containing at a 1:10,000 dilution. The blot was then was washed in TBST three times for 10 min, each, to remove all weakly bound primary antibody. After that, the blot was incubated in the secondary antibody solution at a dilution of 1:20,000, for 1 h at 25 °C. The TBST wash steps

were then repeated. The blot was then incubated in detection buffer (Table 3.2.1) for 10 min, and finally in the developing buffer (Table 3.2.1) and incubated in a dark place without shaking. Bands usually started to appears after a few minutes. The blot was analysed and recorded using the Imagequant Las 4000 system (GE healthcare life sciences, Chicago, USA).

Table 3.2.1 Western Blot Buffers

| Buffers | Composition |
|---------------------------------------|---|
| Blotting buffer | 3.03 g Tris-HCl, 14.4 g Glycine, 200 ml methanol and water up to 1 L. |
| TBS buffer | 24 g Tris-HCl, 5.6 g Tris base, 88 g NaCl water up to 1L (pH 8.0). |
| TBST buffer | 100 mL TBS, 900 mL dH ₂ O, 1 mL Tween 20 . |
| Blocking buffer and antibody solution | 5% (w/v) Bovine Serum Albumin, in TBST |
| Detection buffer | 1 M Tris-HCl, 0.1 M (pH 9.0). |
| Developing buffer | 0.1 CSPD substrate (Invitrogen), 99 ml detection buffer. |

3.2.5. Protein purification

His-tagged proteins were purified using the HisTrap protein purification 1ml column (GE healthcare life sciences GE healthcare life sciences, Chicago, USA, 17-5247-01) depending on the Ni-NTA technology.

The column was first washed with water 5 times its size, before equilibration with an equal volume of lysis buffer. Then, the column was loaded with the crude extract soluble protein. The column was then washed with several washing buffers that comprised different imidazole concentrations (20, 30, 40, 50 and 60 mM), and finally with elution buffer (250 mM imidazole). The amount of washing buffers and the elution buffer that used were 5 times the size of the column size. We had to be so careful to avoid any air getting inside the column throughout the purification process.

The washing outcomes from each washing buffers and the elution buffer were collected in graded test tubes. And a sample of 15 µl from each tube was mixed with 5ul of Tricine Sample Buffer, and loaded onto SDS-PAGE gel. The column was then washed with lysis buffer, water and finally filled with 20% ethanol to be stored at 4 °C. Protein samples were dialyzed to remove all binding imidazole using the 12-14 kDa dialysis tube (Medicell, London, UK), and concentrated using Vivaspin 20 tubes (5 kDa) (GE Healthcare, Buckinghamshire, UK).

3.2.6. *In Vitro* Recombination Assay

Escherichia coli Integration host factor (IHF) is a heterodimeric protein that acts important roles in a variety of cellular mechanisms not only site-specific recombination, but also transcription, and DNA replication. In order to get a better expression of IHF protein, pEE2003 (derivative of pET16b, 5711bp), harboring IHF genes was isolated from our -80°C bacterial stock and retransformed into BL21 *E. coli*. The transformation was tested using both plasmid backbone primers and IHF genes primers. IHF protein overexpression was induced using different IPTG concentrations; purified using a nickel-sepharose column and the overexpression was tested by both SDS-PAGE and semi dry western blot analysis. On the other hand, the $\Phi 24_B$ integrase, cloned into pBAD/Myc-His C plasmid (~4.1 kb), was overexpressed from *E. coli* TOP 10. The overexpression was tested by both SDS-page gel and semi-dry western blot assay.

Additionally, pCR2.1 plasmids (~3.5 kb), harboring a region of DNA encompassing either the *attP* or *attB*, were isolated from *E. coli* TOP 10 strain. In order to remove all small unwanted compounds, like imidazole, both of IHF and Integrase proteins were purified from cell extracts, dialyzed using dialysis membrane tube and concentrated (Rigaut et al., 1999). The protein solution was concentrated using Vivaspin 20 column (GE Health Care). The concentration of each protein was measured using the Pierce BCA Protein Assay Kit.

Several recombination tests, in two replicates for each, were run using either supercoiled *pattP* with supercoiled *pattB*, supercoiled *pattP* with linear *pattB* or linear *pattP* with supercoiled *pattB*. Furthermore, we used *E. coli* cell crude extracts without any protein overexpression. Both *pattP* and *pattB* were cut using the *Bam*H I restriction enzyme to make a linear DNA product.

Approximately 500 ng each of linear or supercoiled substrate DNAs was mixed with 200 ng of approximately 90% pure integrase with and without 100 ng of approximately 90% pure IHF, in addition to about 500 ng of cell extract, in a final reaction volume of 50 μ l (Fig. 3.3).

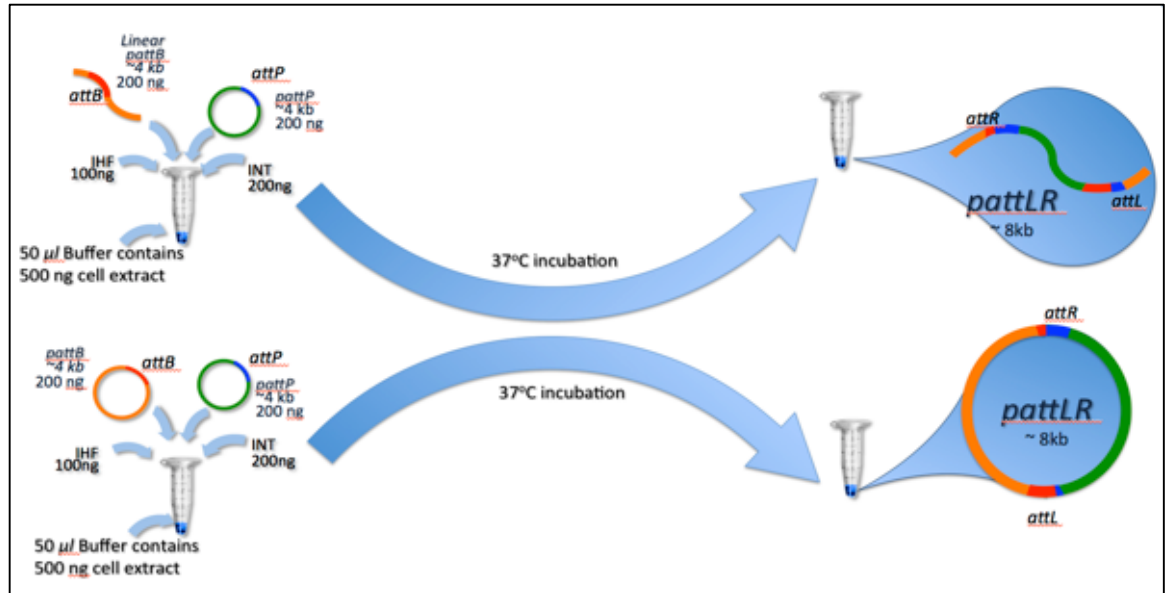


Figure 3.3. *In vitro* recombination assay using *pattB* and *pattP* plasmids and required proteins in a final volume of 50 μ l of suitable buffer. recombination tests were run using either supercoiled *pattP* with supercoiled *pattB* (The bottom test), supercoiled *pattP* with linear *pattB* (The top test) or linear *pattP* with supercoiled *pattB*. In each test the DNA substrate was mixed with Int with or without IHF in addition to cell extract.

Two types of buffer were used to achieve the final reaction condition; the first (Buffer1) one was composed of 20 mM Tris, pH 7.5, 100 mM NaCl, 1% glycerol, 0.1 mM EDTA (Thorpe and Smith, 1998), while the second final condition (Buffer2) was 50 mM Tris-Cl, pH 7.8, 60 mM KCl, 250 mg of BSA for each 15 μ l, 0.5 mM EDTA, 10% glycerol and 5 mM spermidine (Goodman et al., 1999). The reactions were incubated for 6 hours or 24 hour at 37°C, and stopped by separating the recombination solutions in the electrophoresis gel. Any expected recombinant band has been cut from the gel, extracted using gel PCR and Gel kit from Bioline, and used as PCR template using attRL F/R primers.

3.3. Results

The minimal components needed for λ phage *in vitro* recombination, are *attP* and *attB* sequences, purified integrase, and purified *E. coli*-encoded integration host factor. Due to the fact that the Φ 24_B integrase is a member of the tyrosine family of recombinases, like the λ phage integrase, and Φ 24_B is a member of the lambdoid family of phages it was reasonable to assume that the integrase of Φ 24_B would work similarly to the integrase of λ phage. To these ends the various reagents were produced including IHF, *attP* _{Φ 24B}, *attB* _{Φ 24B}, and Int _{Φ 24B} and IHF.

3.3.1 IHF production

The plasmid pEE2003 possessing the *himA* and *himD* genes encoding IHF was kindly donated by E. Haggard-Ljungquist (Table 2.2.1.). In order to utilize IHF in

an *in vitro* integration assay, it was necessary to purify IHF. Using IHF from plasmid pEE2003-IHF where the α subunit is labelled with a molecular tag consisting of 6 contiguous histidine residues, both subunits, forming a heterodimer, could be purified by Ni-affinity chromatography.

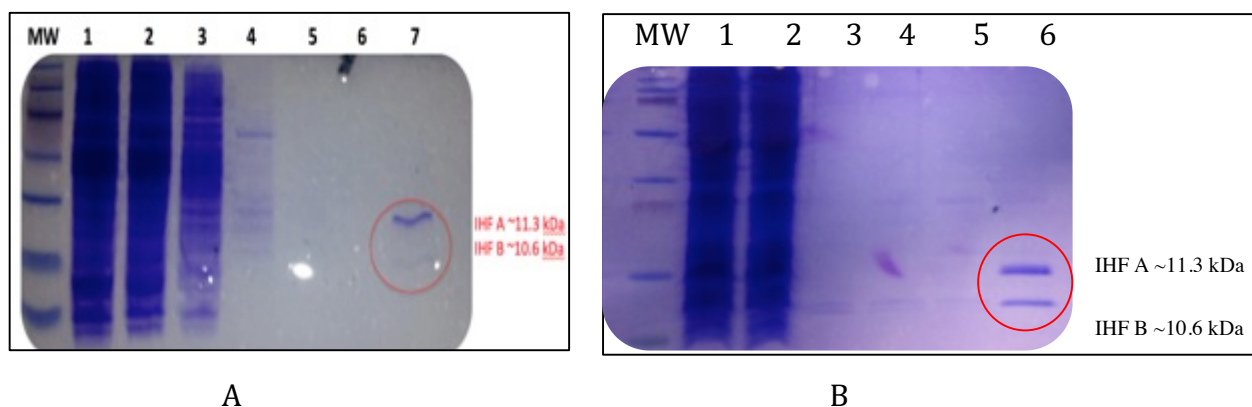


Figure 3.4: Production of IHF from a His-tagged cloned (kindly provided by E. Haggard-Ljungquist). The low expression of IHF(A) was improved when its pEE2003-Ihf plasmid was isolated and retransformed to new chemically competent cells for protein overexpression and purification. We can see in picture B, SDS-page gel of His-tagged IHF purification steps. MW, prestained molecular weight markers (2-250 kD) (BioRad); 1, whole cell lysate; 2, Ni-column flow through; 3-5, column washes; 6, IHF in 60 mM imidazole washing buffer.

Conditions for induction of IHF expression were tested (Fig. 3.4A) and optimized following the isolation and retransformation of pEE2003-IHF into freshly prepared chemically competent cells (See 2.12. Transformation of chemically competent cells)(Fig. 3.3B).

3.3.2 Int_{Φ24B} Production

The plasmid pBAD/Myc-His C, possessing the *int* gene encoding Int was kindly donated by P. Fogg (Table 2.2.1.). In order to utilize Int in an *in vitro* integration assay, it was necessary to purify Int, the main protein that performs DNA recombination. This was done using Int from plasmid pΦ24B-int where the protein molecule is labelled with a molecular tag consisting of 6 contiguous histidine residues, and could be purified by Ni-affinity. The conditions for induction of Int were tested (Fig 3.5).

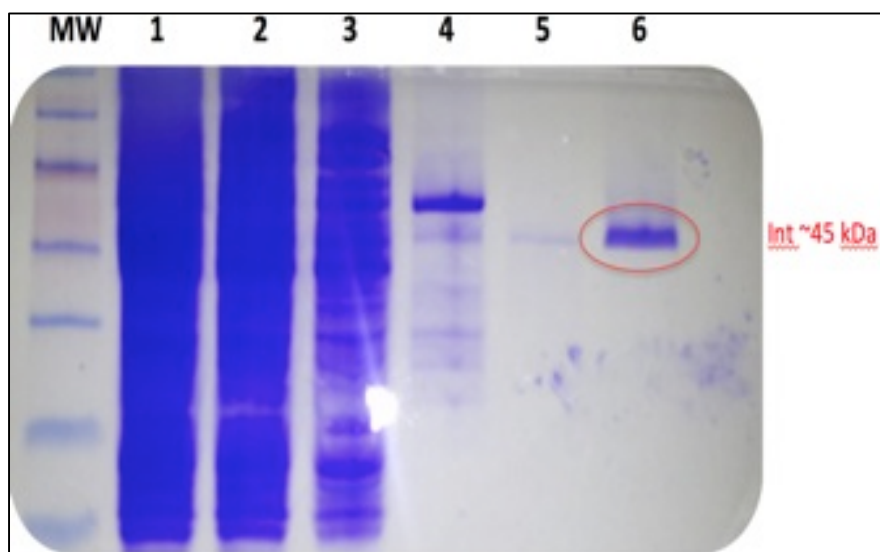


Figure 3.5. Production of Integrase from a His-tagged cloned. SDS-page gel of His-tagged Integrase purification steps. MW, prestained molecular weight markers (10-250 kD) (BioRad); 1, whole cell lysate; 2, Ni-column flow through; 3-5, column washes; 6, Integrase in 60 mM imidazole.

3.3.3 Production of attP Φ 24_B, attB Φ 24_B constructs

In order to establish a working protocol for an *in vitro* integration assay, we need to insure that we have all components needed for this *in vitro* recombination to be performed by the act of Int Φ 24_B and IHF. The DNA substances for these proteins are *attP* Φ 24_B, and *attB* Φ 24_B. Therefore, different sizes of both *attP* and *attB* sequences were amplified (See 2.5 PCR, Chapter 2) and cloned into their appropriate plasmids (Table 2.2.1).

Plasmids harboring either the 100 bp, 200 bp, 300 bp, 400 bp or 600 bp *attB* sequences were confirmed using the *attB* 50 F/R *attB*100 F/R, *attB*150 F/R, *attB*200 F/R and *attB*300 F/R primers, respectively, as well as the M13 F/R and primers (The plasmid backbone primers). Further more, the transformation of 100 bp, 200 bp, 300 bp, 400 bp or 600 bp *attP*-cloned pCR2.1 plasmids, were confirmed using *attP* 50 F/R *attP*100 F/R, *attP*150 F/R, *attP*200 F/R and *attP*300 F/R primers, respectively, as well as the M13 F/R and primers (Fig. 3.6 & 3.7).

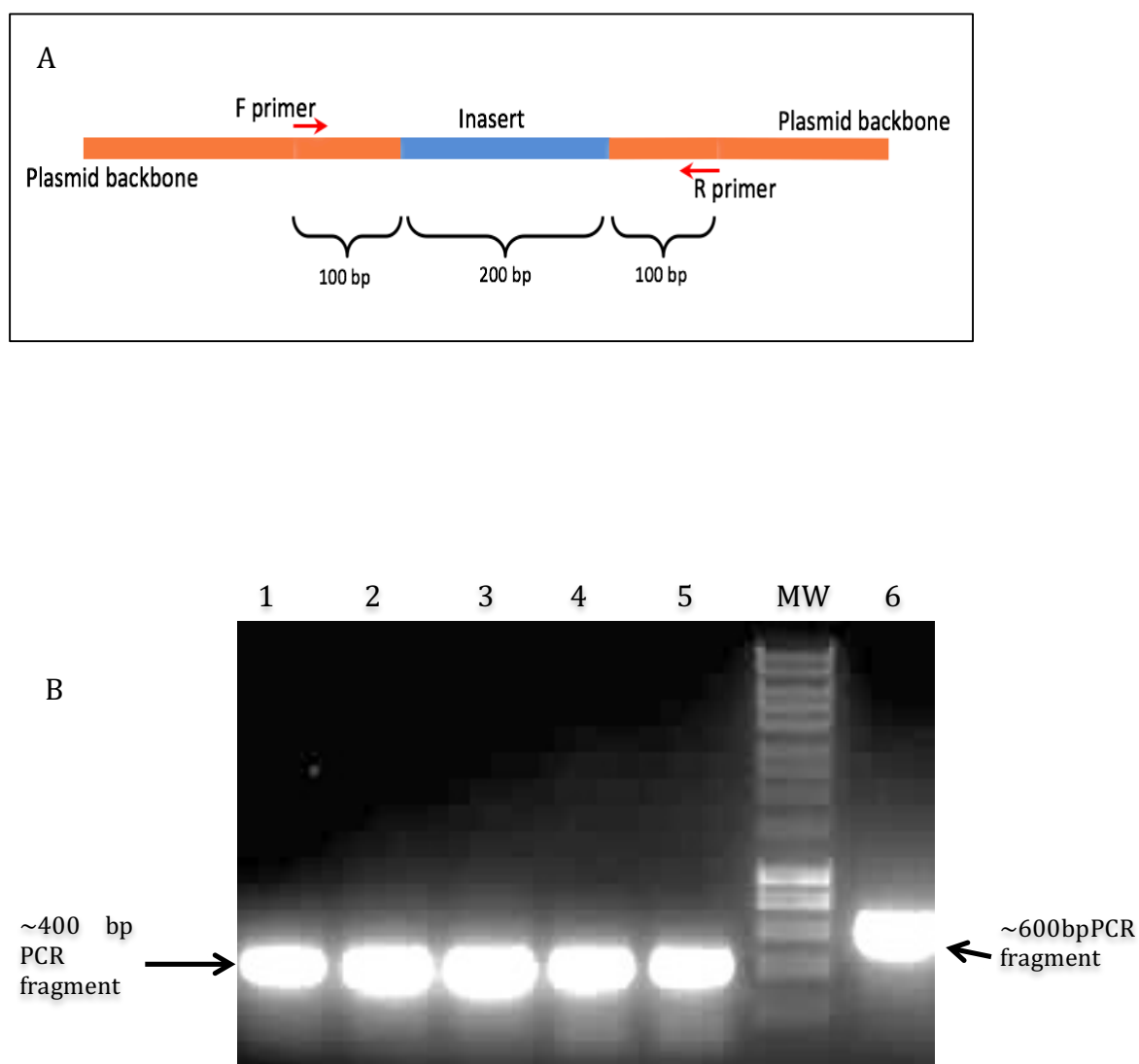


Figure 3.6. Colony PCR of the *E. coli* Top 10 transformants to confirm the construction of pPCR2.1-attB200 bp. **A.** schematic diagram showing the sites of the backbone primers (Red color) that used in the PCR reaction, the PCR products were generated with the M13 F/R primers. **B.** The electrophoresis agarose gel picture 1-5, Colony PCR ~400 bp fragments amplified using M13 F/R primers.; MW, 1kb Hyperladder I (Bioline); 6, positive control using plasmid with *attB* 600 bp insert as template.

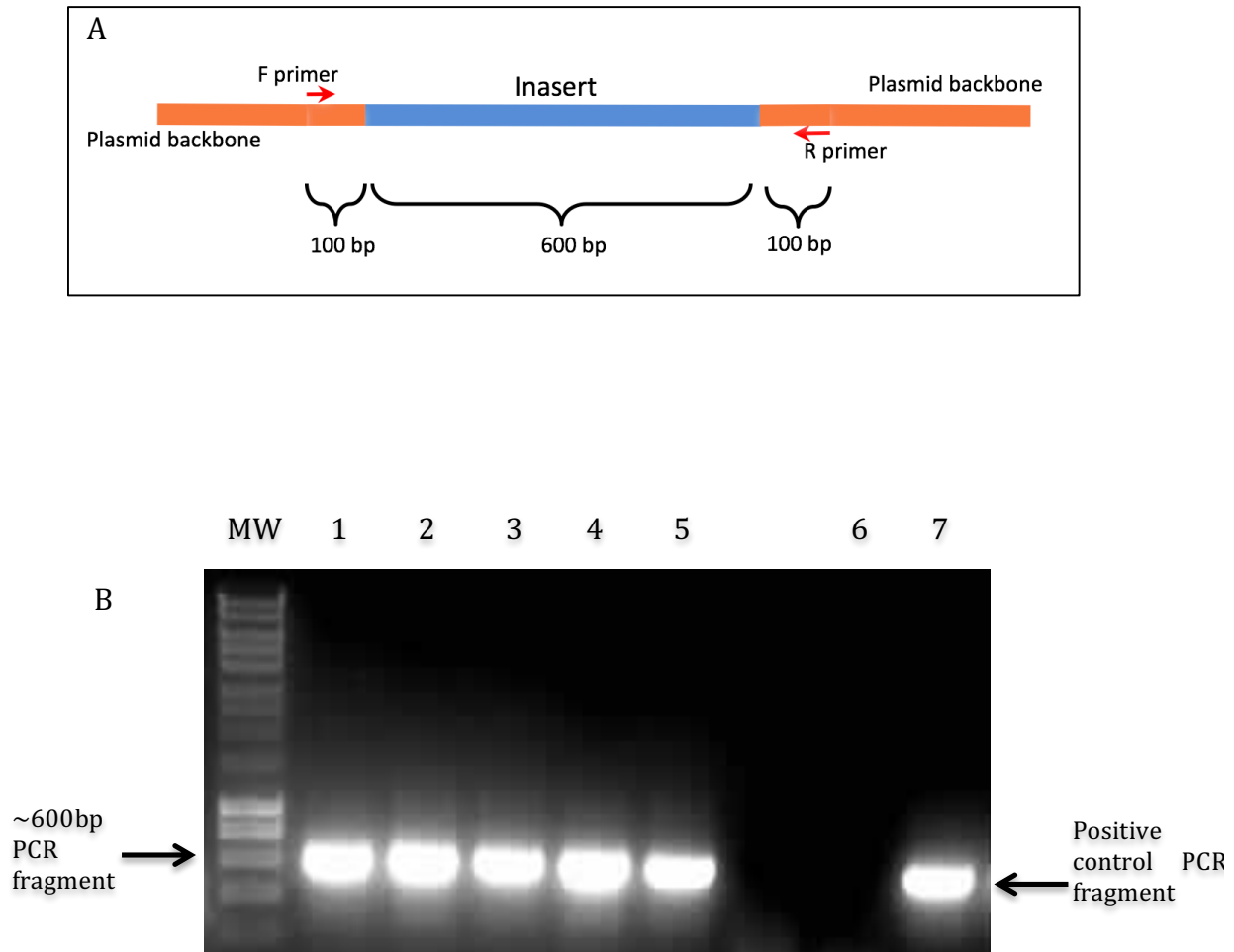


Figure 3.7. Colony PCR of the *E. coli* Top 10 transformants to confirm the construction of pPCR2.1-*attP*400 bp. **A.** schematic diagram showing the sites of the backbone primers (Red color) that used in the PCR reaction, the PCR products were generated with the M13 F/R primers. **B.** The electrophoresis agarose gel picture: MW, 1kb Hyperladder I (Bioline); 1-5, Colony PCR ~600 bp fragments amplified using M13 F/R primers.; 6, Negative control; 7, positive control using plasmid with *attB* 400 bp insert as template.

3.3.5 Integration Assay (*in vitro*)

Before the first attempt at the integration assay, the presence of unwanted nucleases in the recombinant Int_{Φ24B} and IHF preparations was examined (Fig. 3.8).

To analyze, if Φ24_B integrase is able to promote recombination between *attP* and *attB*, *in vitro* recombination assay was done with present of both IHF and Int proteins. This assay was repeated more than one time using different buffers in addition to Top 10 cells crude extract without any gene overexpression.

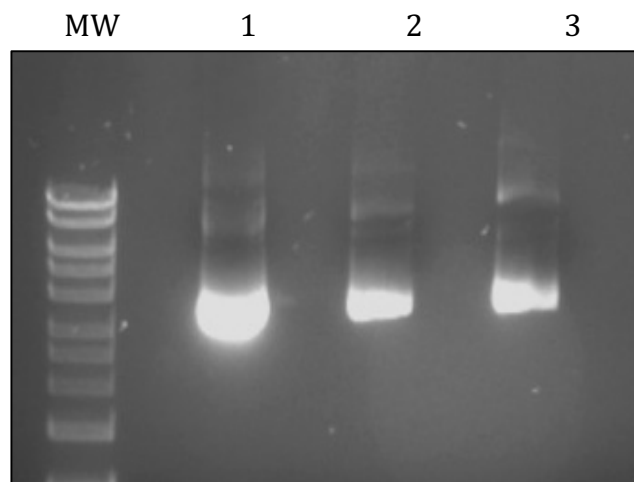


Figure 3.8. Quality control for purified IHF and Integrase. MW, 1 kb DNA molecular weight marker; Lane1, plasmid (100 ng) carrying *attP* composed of a 600 bp DNA segment (300 bp flanking either the overlap region) as control. Same plasmid was incubated with 1 µg of either IHF (lane 2) or Integrase (lane3) at 30 °C for 1 hour. No co-purifying nuclease activity was detected.

On 0.7 % of agarose gel no recombination could be detected in any *in vitro* assays , with exception of those run with crude cell extract with linear *pattB* and supercoiled *pattP* (Fig. 3.9). This positive result was confirmed by harvesting the expected recombinant band from the gel, extracted using gel PCR and Gel kit from Bioline, and using PCR and the primer pairs M13 F/R or attRL F/R (attP150F/attBR) to confirm the identity of the integrand (Fig.3.10).

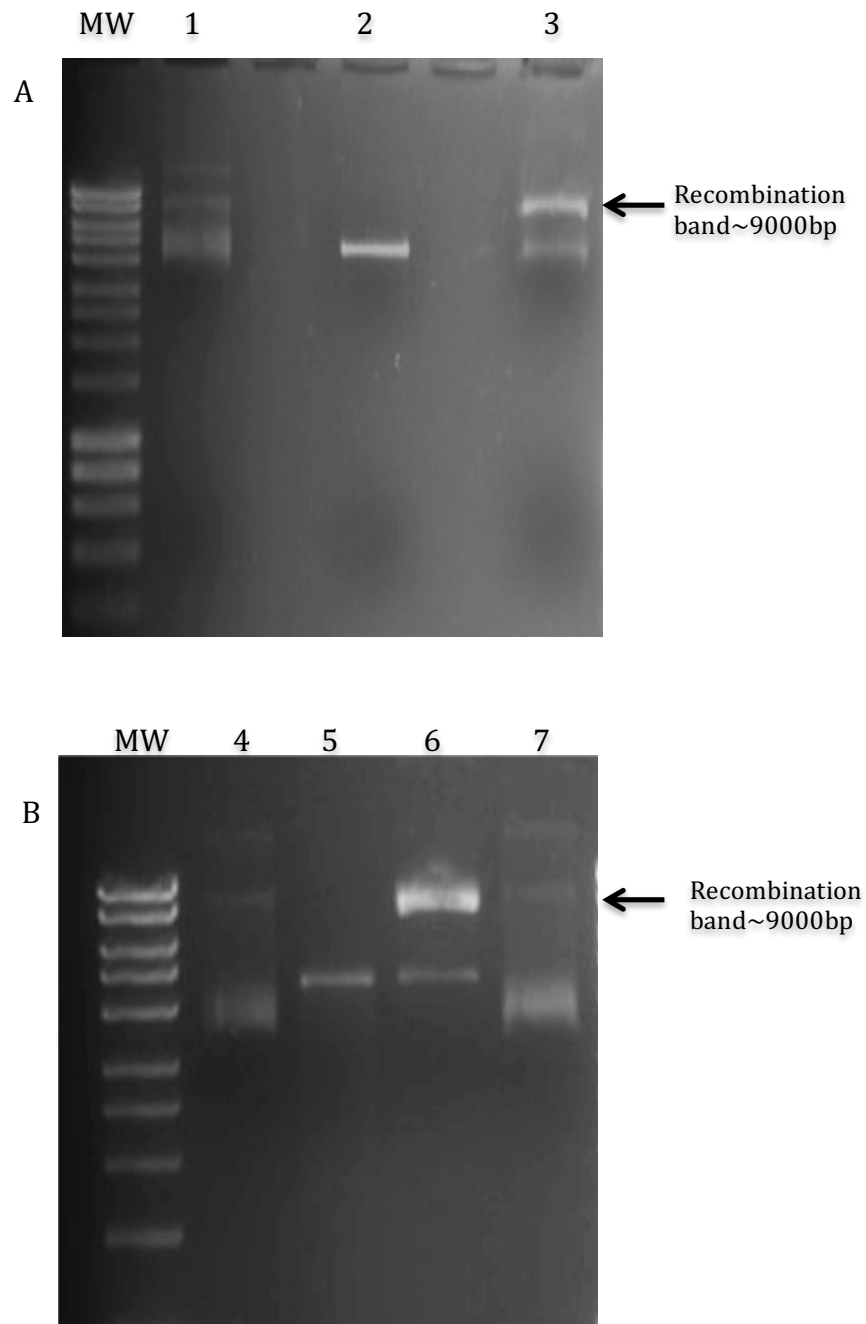


Figure 3.9. *In vitro* recombination tests using cells crude extract. A, After 6hr. at 37°C incubation. **B,** after 24hr. at 37°C incubation; 1,4, test tubes with coiled *pattP* only without cell extract (negative control); 2,5, test tubes with linear *pattB* only without cell extract (negative control); 3,6, recombination test with Linear *pattB* and supercoiled *pattP* in addition of cell extract;7, supercoiled *pattP* and Linear *pattB* without any other factor as negative control .

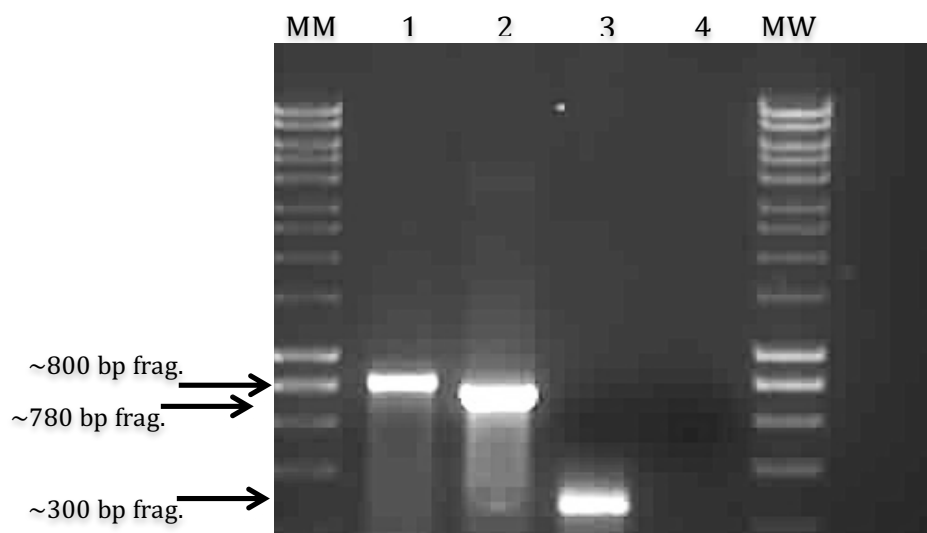


Figure 3.10. PCR assay of the recombination products. MW, 1kb Hyperladder I (Bioline); 1, positive control using M13 F/R primer with pCR2.1-*attP600* as template; 2, *attRL* PCR fragment amplified using M13 F/R primers; 3, *attRL* PCR fragment amplified using *attP150* F/*attB* R primers.

3.4. Discussion

Previous studies have demonstrated that experimental parameters required for various integrases in *in vitro* recombination assays are not always the same. In case of recombination catalyzed by the bacteriophage P2-encoded integrase, a tyrosine recombinase, integration requires 2 *att* sequences, integrase and IHF (Sylwan et al., 2010), a phenomenon not uncommon across tyrosine recombinases (Groth and Calos, 2004). However, there are tyrosine recombinases like Cre, encoded by bacteriophage P1 that function without accessory proteins (Groth and Calos, 2004). The Φ C31 recombinase is of the resolvase/invertase family and can catalyze reactions in the absence of integration host factor. This reaction was

measured over a time period of more than 8 hr in *in vitro* assays (Thorpe and Smith, 1998).

As $\Phi 24_B$ integrase belongs to the tyrosine family recombinases and most of the members of this family need the accessory protein IHF to drive recombination, we considered IHF would be an important factor for $\Phi 24_B$ integrase to promote recombination. However, we did not obtain any recombination product unless we added crude *E. coli* cell extract to the reaction. For this we have to mention that, Sylwan *et al.*, (2010) reported the *in vitro* recombination between bacteriophage P2 attachment site and *E. coli* attachment site attB, can be catalyzed by adding a mix of two crude extracts, one of them is overexpressing P2-encoded integrase and the other overexpressing IHF protein. However, in our case, having positive results after adding crude *E. coli* cell extract means one of two things: 1) the IHF purified from pEE2003-IHF was not biologically active, or 2) there is another bacterial host accessory factor(s) needed for the $\Phi 24_B$ integrase to function as a site-specific recombinase.

Moreover, recombination was only detected in the *in vitro* assay when *pattB* was linearized, a phenomenon reported by Frumerie *et al.*, (2008), who demonstrated that the use of linear *attB* DNA fragments with supercoiled *attP* plasmid and IHF was necessary to facilitate recombination by the P2 integrase enzyme *in vitro*. This assay was cumbersome to perform and resulted in only a limited amount of success. In order to be able to better characterise the $\Phi 24_B$ integrase the decision

was made to move on to an *in situ* assay (Fogg et al., 2011). This assay does not allow the parameters of the recombination events to be as strictly controlled, but has been shown to be dependable/robust in the detection of recombination.

CHAPTER 4. *IN SITU* RECOMBINATION ASSAY

4.1. Background

Since the first identification of a site-specific recombination in the integration and excision of bacteriophage lambda genome from a specific site in the *E. coli* genome (Landy, 2015), many site-specific recombination systems have been identified and characterized (Groth and Calos, 2004). Most of these recombinases have been associated with bacteria, but some have also been characterised from archaea and eukaryotic cells, like yeasts (Jayaram et al., 2015; Smith and Thorpe, 2002). DNA rearrangement by site-specific recombinases can be performed by bringing two DNA strands together. These DNA sites can be either at separate locations within a single DNA molecule or just simply located in two different molecules. Both IHF and integrase proteins then work by protein-protein interaction and DNA homology. At a specific point within the sites, the DNA strands are broken and reorganized in strand exchange reaction, which gives new DNA product (Fig 1.10, Chapter 1) (Craig and Nash, 1984). The result of this reaction can be integration, inversion or excision of the DNA substrate. The reaction outcome and duration are precisely controlled. In case of temperate phage, this strict control is important to minimize the opportunity for DNA damage during the rearrangement. It is essential for phage genome integration and later phage excision and for lytic or lysogenic life cycle (Smith and Thorpe, 2002).

In λ phage, the archetypal phage for the lambda family, the phage genome, at entry into the lysogenic replication cycle, integrates inside the *E. coli* genome after

the recognition of the specific attachment site in the bacterial chromosome, the *attB* site. This site recombines with the the phage attachment site *attP*. The outcomes of this process are two hybrid sites, *attR* and *attL*, flanking an integrated prophage genome within the host chromosome. Each new hybrid site is comprised of about half of the *attB* site and half of the *attP* site (Ross et al., 1979).

Driving the incorporation of the λ prophage into the now lysogenic host is the phage-encoded integrase (Int) enzyme, which is a member of the tyrosine family of recombinases (Enquist et al., 1979). The specificity of recombination sites is determined by shared sequence within the *attB* and *attP* sites, and the set up for recombination is driven by both DNA-DNA sequence homology and DNA-protein interactions (Weisberg et al., 1983). As such, λ phage integrase interacts with the specific sites, *attP* and *attB*, producing an Int/DNA intasome complex, bringing the sites together to catalyze a unidirectional reaction that produces and then resolves a Holliday junction intermediate, which ultimately results in the production of an integrated prophage (Chen et al., 2000; Azaro and Landy, 1997)

. Homology between overlap regions is very necessary to promote recombination (Echols and Green, 1979) (see also 3.1.2 Lambda phage integrase, Chapter 3).

To mediate the DNA recombination, λ phage integrase needs an *attB* site consisting of 21 nucleotides with an integration overlap region, also known as the

point of crossover (O_B), whereas. The *attP* site is 250 bp long and possesses a crossover region (O_P). The lambda overlap region is 7 bp (Fig.1.10. Chapter 1).

A bacterially encoded protein, integration host factor (IHF), is also required to facilitate the integration process (Sylwan et al., 2010, Groth and Calos, 2004). It has been reported that IHF is essential for λ phage site-specific recombination. To facilitate integrase activity, IHF binds to regions in *attP* at specific arm binding sites, enabling integrase to bind to the *attP* site-arms with high affinity while maintaining a lower affinity to the *attP* core-binding region (Richet et al., 1988), (see also 3.1.2 Lambda phage integrase, Chapter 3). The recombination (Fig.4.1) requires each monomer to have an N-terminal domain that resides over each adjacent core-binding region (Biswas et al., 2005). The C-terminal domains (core binding and catalytic domains) recognize and bind to the identical crossover sequences in the *attP/attB* regions in a square planar formation, promoting DNA recombination (Fig. 4.1) (Van Duyne, 2005).

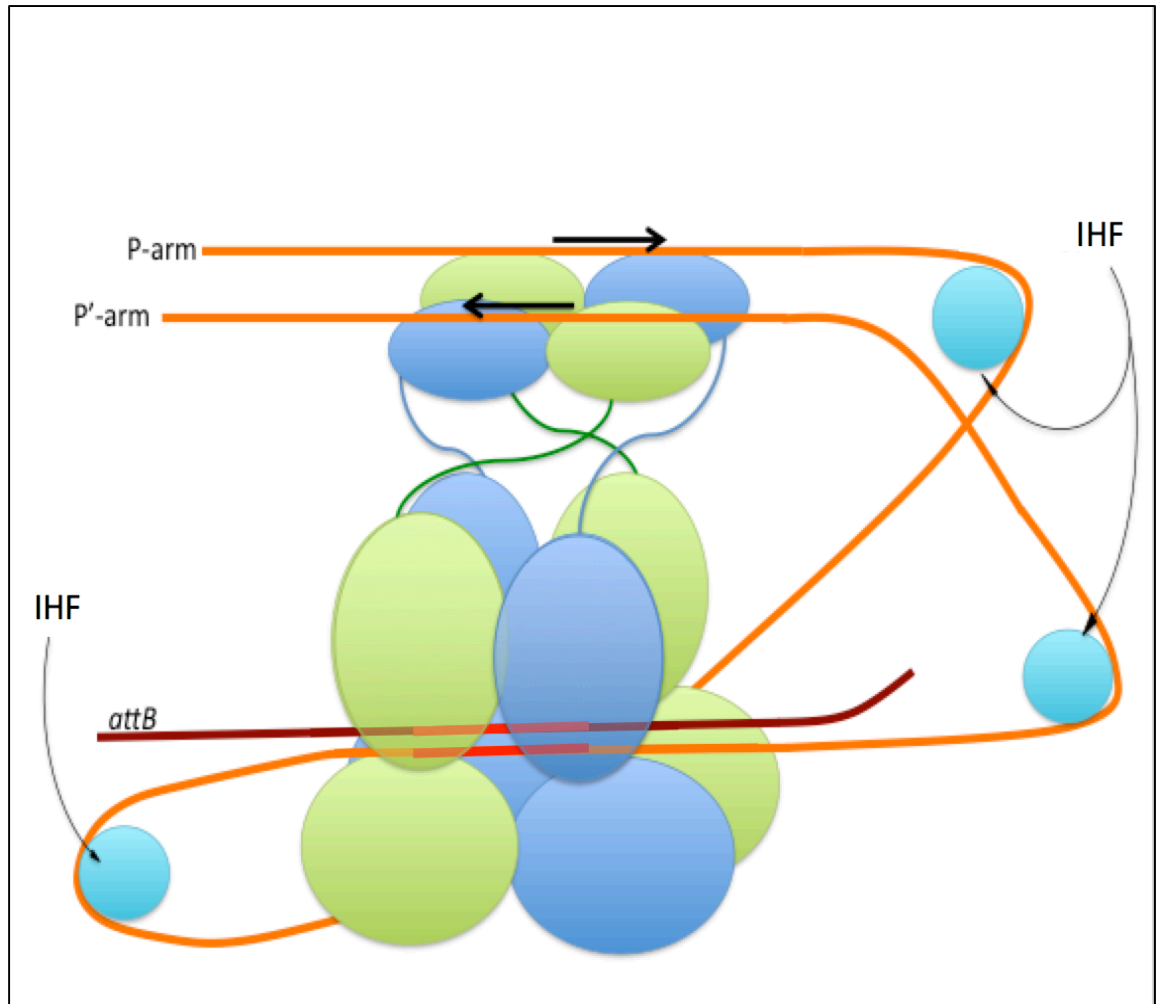


Figure 4.1. The Int/DNA intasome complex based on the models proposed by (Biswas et al., 2005). A cartoon depiction of the protein-DNA interaction during bacteriophage λ integration. The integrase protein binds with high affinity to the *attP* site-arm regions and with low affinity to the *attP* core-binding region. Integration host factor is necessary for extensive *attP* DNA binding that required for intasome formation. Red colour lines highlight the crossover region.

4.1.1. The Stx Phage $\Phi 24_B$

$\Phi 24_B$ is an Stx encoding bacteriophage that is a member of the lambdoid family (Allison, 2003). All members of this family have a double stranded DNA genome, and have similar morphological characteristics (icosahedral head with tail). Moreover, they are genetically very similar. Lambda phage λ , the archetype of this family, possesses a genome of 48.5 kb, which is enough to govern all its biological aspects (Biswas et al., 2005). $\Phi 24_B$ has a larger genome (57.67 Kb) than lambda. Little is known about the function of this extra DNA carried by $\Phi 24_B$ (Allison, 2003).

Bacterial and phage integration attachment sites were identified by the alignment of the $\Phi 24_B$ genome sequence (*attP*), the MC1061 genome sequence (*attB*) and the integrated prophage (within lysogen genome) sequence (*attL*) (Fig.4.2) (Fogg et al., 2007). The primary bacterial integration site is located within the *E. coli* genome, just upstream to the integrase gene within the cryptic prophage CPS-53.

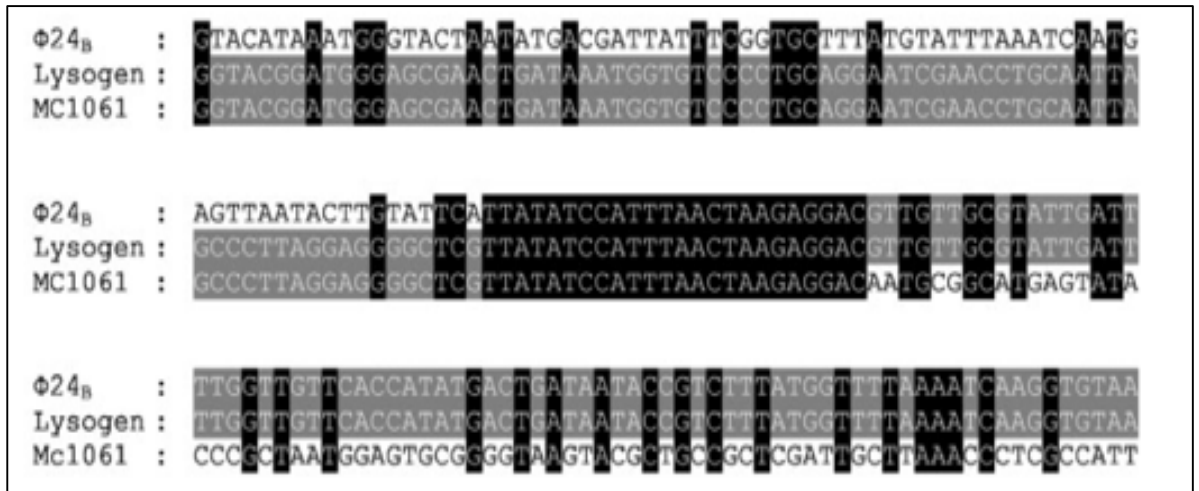


Figure 4.2. The identification of bacterial and bacteriophage integration attachment sites. The following sequences were aligned: Φ24_B genome sequence (*attP*), the MC1061 genome sequence (*attB*) and the integrated prophage (within lysogen genome) sequence (*attL*) (Fogg et al., 2007) to identify the core cross over regions.

4.1.2. *In situ* assay

The term *in situ* is a Latin phrase that means “on site” or “locally”, and it describes the scientific measurement event that happens at the same place as the original phenomenon is naturally occurring without altering the original conditions (Lewis et al., 1969). In biology and biomedical engineering, *in situ* means to examine the phenomenon exactly in place where it occurs.

We call the system that we used for the intracellular plasmid recombination an *in situ* recombination system, as we run the assay at the same original conditions that $\Phi 24_B$ Int needs in order to facilitate DNA recombination between the $\Phi 24_B$ phage genome and the *E. coli* bacterial genome.

Previously, we were trying to develop a repeatable system in which we could precisely control all of the components involved in the recombination reaction (DNA substrates, enzymes and proteins), so we started with an *in vitro* recombination system. where we could control the quantities and ratios of DNA, enzymes, and proteins as well as the reaction conditions. However, *in vitro* assays require that all of the required factors are known, and, our limited *in vitro* assay results led us to wonder if there was either an unknown, *E. coli* encoded protein necessary for $\Phi 24_B$ Int activity, or if the purified IHF used in the *in vitro* assay lacked biological activity. Therefore, we moved to work with a system that would provide the same set of conditions that occur within the host cell, where we knew we could detect Int-mediated recombination. This system was the three plasmids *in situ* recombination assay.

In order to run an efficient *in situ* plasmid system, we have to avoid any plasmid incompatibility events. Plasmid incompatibility can be defined as the failure of two, or more, co-resident plasmids to persist in a cell. In other words, if transformation of second plasmid negatively affects the inheritance of the already

existing first plasmid, the two plasmids are considered to be incompatible. The reason behind that is the plasmid's origin of replications (*ori*). It is impossible to have two different plasmids that use same mechanisms of replication in one single cell (Novick et al., 1976). Therefore, to maintain three plasmids in an *in situ* system, we had to use compatible plasmids, and that means each plasmid has its own replication strategy and different from the other two.

4.1.3 Aims

- Cloning of both *attB* and *attP* into separate compatible plasmids to run recombination between them.
- Transformation of both *att* site plasmids and pΦ24B-into *E coli* competent cells (one cell harbours 3 plasmids).
- Run DNA recombination driven by Φ24B-INT protein using *in situ* recombination assay.
- Using the *in situ* recombination assay, determine the minimal sizes of both *attB* and *attP* sites that are necessary to support integration.
- Identify the necessary size of each arm binding sites (B and B') of *attB* recognised by the integrase and surrounding the 24 bp overlap region or the core-binding site (O).
- Identify the importance of the bacterial-encoded, integration host factor for integrase activity.

4.2. Specific Methods

4.2.1. Plasmids cloning of *attB* and *attP* sites

An already prepared pCR- Φ 24B-attP600 plasmid (Table 2.2.1.), harboring the attP core and 600 bp of flanking sequence (see 3.2.1. Cloning of bacterial and phage attachment sites, Chapter 3), was isolated from *E. coli* TOP 10 strain (Invitrogen). The *attP* fragment was cut using *Bam*H I and *Xho*I endonucleases enzymes. The same digestion enzymes were used to cut the high copy number compatible plasmid, pCDF-Duet. After the endonucleases were deactivated, each of the above was purified by the agarose gel and recovered using PCR and Gel extraction kit from (Bioline, London, U.K.). The *attP* sequence was ligated into pCDF-Duet to form p Φ 24_B-attP₆₀₀ (Table 2.2.1.), before it was transformed to new Top 10 competent *E. coli* cells. At the same time, both of already prepared pCR- Φ 24_B-attB₆₀₀ harboring the *attB* core and 600 bp of flanking sequence, and the medium copy number pACYC-Duet compatible plasmid were cut using same endonuclease enzymes (*Bam*H I and *Xho* I). After the restriction enzymes were deactivated, both of these DNA sequences were agarose gel purified and recovered. The linear pACYC-Duet plasmid was ligated by *attB* sequence to form p Φ 24_B-attB₆₀₀ (Table 2.2.1.), before it was transformed into new Top 10 competent cells (Invitrogen) (Table 2.1.1.).

4.2.2. Restriction free cloning

This method has been described by Bryksin and Matsumura (2010) and (Bryksin and Matsumura, 2010, van den Ent and Löwe, 2006). All *attB* site inserts were amplified from genomic DNA of the *E. coli* K-12 strain, MC1061 (See 2.1. Bacterial Strains, Chapter 2), while, all *attP* inserts were amplified from the DNA of the Φ 24_B phage genome. The PCR amplification of the insert DNA was performed using Q5 hot start High-Fidelity DNA polymerase (New England Biolabs, Ipswich, USA) (See 2.5 PCR, Chapter 2). Primers that were used in these PCR amplifications are listed in Table 2.5.2. (Chapter 2). The forward primer had 20 bp of sequence at its 5' end that was complementary to the cloning site in the plasmid vector, with the last 15-20 bp of the primer being complementary to the insert of interest. The reverse primer was also capable of annealing to the plasmid cloning vector with 20 bp of complementary sequence whilst the rest of the primer was complementary to the insert being amplified and cloned. The total volume of the amplification reaction was 50 μ l. PCR products were run in an agarose gel to confirm the apparent molecular weight and in order to purify the correct sized products using the PCR and Gel extraction kit from (Bioline, London, U.K.).

Each amplified insert was now flanked by DNA sequences homologous to the area of the plasmid vector where the insert was to be cloned. After the gel extraction of the PCR product, for each cloning, 200-300 ng of insert PCR product was used in the linear amplification reaction with 20-30 ng of the desired plasmid, 200 μ M dNTP's (Bioline, London, U.K.), 1 \times of Q5 reaction buffer, and 0.02 U/ μ l of Q5 hot

start High-Fidelity DNA polymerase. The total volume of the reaction was 50 μ l. By denaturing the insert and plasmid DNA the homologous sequences shared between insert and plasmid vector could align and the insert DNA was now free to prime the extension of the plasmid vector in a linear amplification reaction. The reaction was heated for 30 s at 98 °C, followed 25 cycles composed of denaturation at 98 °C for 10 s, annealing at 60 °C for 30 s, and extension at 72 °C for 30 s/kb and final extension for 5 minutes at 72 °C. The resulting product was the plasmid vector with the cloned insert sequence.

After the PCR reaction was completed, the PCR product was purified using the PCR and Gel extraction kit from (Bioline, London, U.K.). The DNA was eluted with 10 μ l of pre-heated water (50°C) before the spin column was centrifuged at 11,000 g for 1 min, as the finale step of that purification (See 2.6.1 Recovery of the DNA from the agarose gel, Chapter 2).

In order to digest the methylated parental plasmids, 8.5 μ l of the restriction free cloning reaction were treated with 0.5 μ l of *Dpn* I (Bioline), and 1 μ l of CutSmart Buffer (Bioline, # B7204S) for 2 h at 37 °C. Then, 5 μ l from the reaction was transformed into new Top 10 Competence cells and transformants were selected for on LB agar with the plasmid resistance antibiotic at its specific concentration (Invitrogen)(Table 2.1.1.). Positive cloning was confirmed by plasmid prep and DNA sequencing (Fig. 4.3).

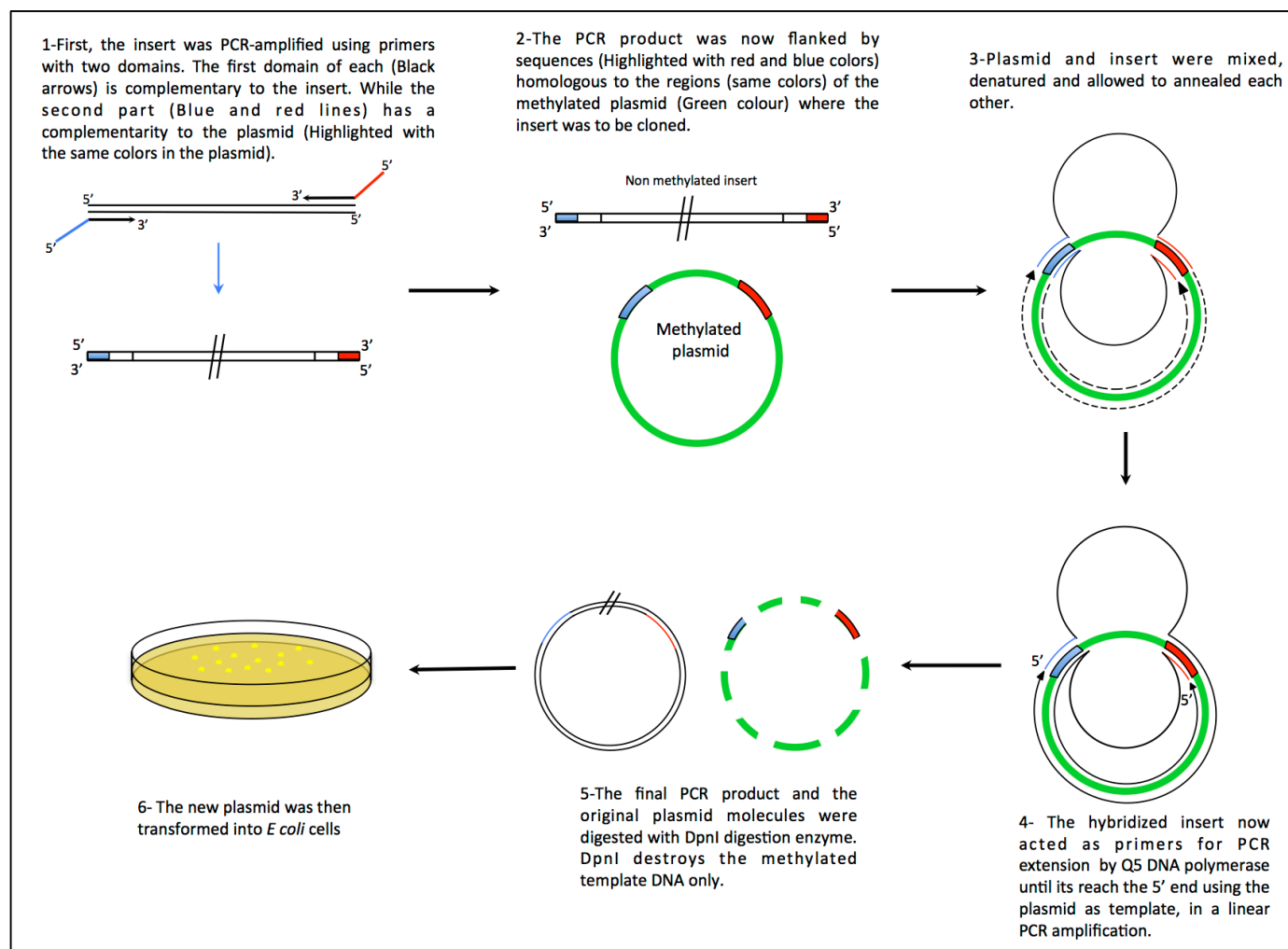


Figure 4.3. Restriction free cloning (Adapted From Bryksin and Matsumura, 2010). Using primers comprised of two distinct sequence domains (insert and plasmid sequences) the insert to be cloned was amplified. This amplification product was then flanked by sequences homologous to the insertion site in the plasmid cloning vector. Using methylated plasmid DNA and the amplified insert product a second amplification reaction was run using the insert DNA to prime the production of the desired DNA clone. Once these primers anneal to their sites in the plasmids, the Q5 polymerase extends and incorporates the insert into a nicked, circular DNA molecule using the plasmid as template until it reaches 5' end of the insert. The parental vector was digested away using the restriction endonuclease Dpn I which cleaves only methylated DNA (this destroying only plasmid DNA in which no insert has been added. The remaining DNA product was transformed into New Top 10 Competent cells.

4.2.3. *In situ* recombination assay

All *in situ* recombination (RL) assays were performed inside One Short TOP10 *E coli* cells (Table 2.1.1.). Each required plasmid transformed sequentially into these cells until all required, compatible plasmids for each assay were in one *E coli* host cell. An appropriate amount of LB broth, with all plasmid's resistance antibiotics at their specific concentrations, was inoculated with 1 ml of an overnight culture of cells harbouring the compatible plasmids. This culture was incubated at 37 °C with shaking at 200 r.p.m til the OD₆₀₀ 0.5-0.8 was reached. Expression of the integrase protein (Int) was induced by the addition of L-Arabinose to the final concentration of 0.02% w/v. The cultures were incubated for an additional three hours at same incubation temperature. Finally, from each assay flask, 5 ml samples were taken and all plasmid DNA present was extracted from the cells in the samples. The assays results were determined following: 1) a PCR reaction using attRL F/R primers, 2) agarose gel electrophoresis and 3) Sanger sequencing of the PCR products (See 2.15. The Sequencing analysis, Chapter 2). All these assays where performed along with negative control, comprised of a sample that simply was not induced with arabinose so that no Int was produced (Fig.4.4).

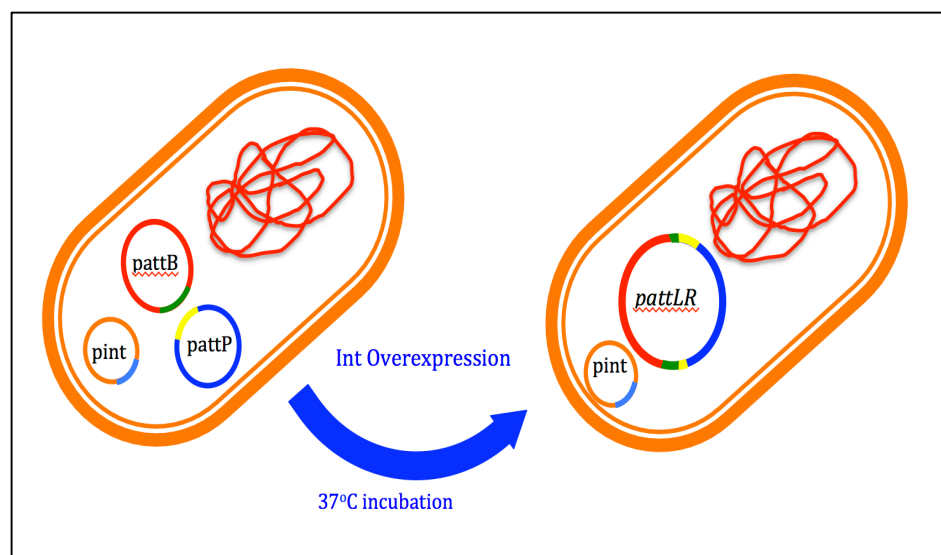


Figure 4.4. *In situ* recombination assay. The ToP 10 *E. coli* cells harboring all of pΦ24_B-attB, pΦ24_B-attP and pΦ24_B-int, were incubated until the right O.D. when the overexpression of Int protein was induced. Int protein in turn drives the recombination between pΦ24_B-attB and pΦ24_B-attP in hybrid recombination product.

4.2.4. The importance of IHF for Φ24B integration

An *E. coli* strain that carries a knockout mutation of the *ihfA* gene was used as a reaction vessel cell in the *in situ* recombination assay. Chemically competent cells were prepared for this strain (See 2.11. Preparation of chemically competent cells, Chapter 2). After the transformation of required each plasmid, successively, all of the following plasmids, pΦ24_B-attB, pΦ24_B-attP and pΦ24_B-int, were introduced into one *E. coli* host cell. The *in situ* assay was run as described above.

4.3. Results

4.3.2. The minimal necessary *attP* size

In order to determine the minimal size of *attP* sequence capable of supporting integration, *E. coli* cells harboring pΦ24_B-attB₁₋₆₀₀ and pΦ24_B-int were used as the host to test different *attP* sequence sizes in separate *in situ* recombination assays (Fig.4.5). For these assays, restriction free cloning was used to clone the site *attP* of either of 140 bp, 230 bp, 350 bp, 427 bp (all possessing the 24 bp site core at their centre), separately, into the high copy number pCDF-Duet. Meanwhile, the 600 bp *attP* sequence was also cloned into the same plasmid using standard restriction enzymes and ligase-based cloning.

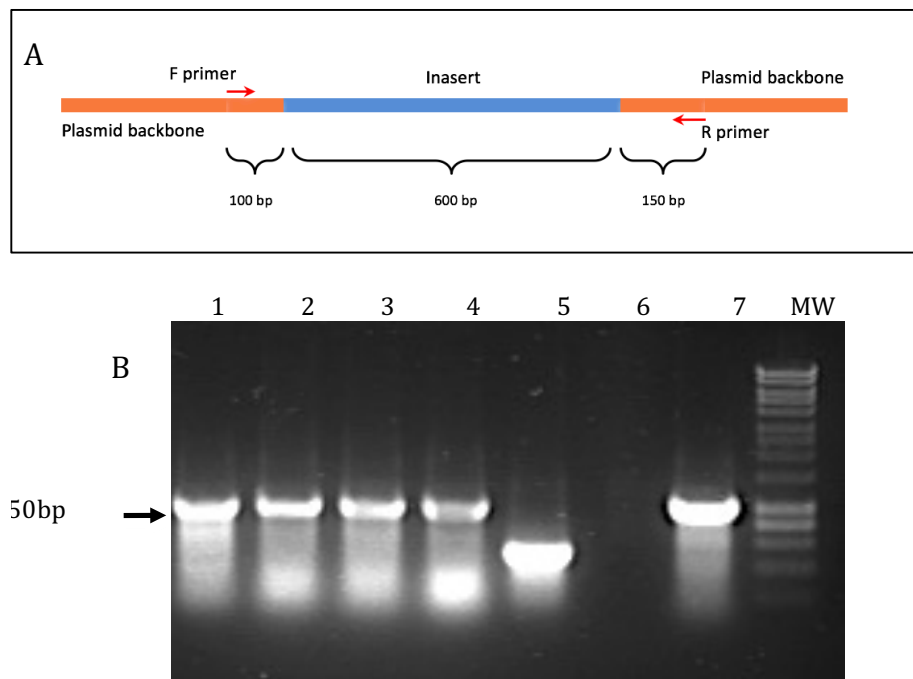


Figure 4.5. Colony PCR of plasmids in the *E. coli* Top 10 transformants confirming the transformation of *pattP* (600 bp) to cells already harbouring pΦ24_B-int and *pattB*. A. A schematic diagram showing the sites of the backbone primers (Red color) that used in the PCR reaction, the PCR products were generated with the Forward ACYCDuetUP1 and Reverse T7 Terminator; B. The electrophoresis agarose gel picture. MW, 1kb Hyperladder I (Bioline); Lanes 1-4&7, ~850 bp bands from 5 individual clones amplified by PCR; lane 5, Positive control: plasmid with 200 bp insert; Lane 6, primers negative control: PCR without template.

Following each *in situ* recombination assays, which were performed in two replicates, integration events were identified by determining the existence of the large hybrid p Φ 24_B-*attRL* plasmid. Plasmids were harvested following the *in situ* assay and PCR-based DNA amplifications were performed using *attRL* primers. Resulting products were analysed on an agarose gel and the identities of the products confirmed by Sanger sequencing. The minimal *attP* sequence necessary to support phage genome integration inside *E. coli* genome was determined to be 427 bp, with almost equal P and P' sizes (Fig 4.6).

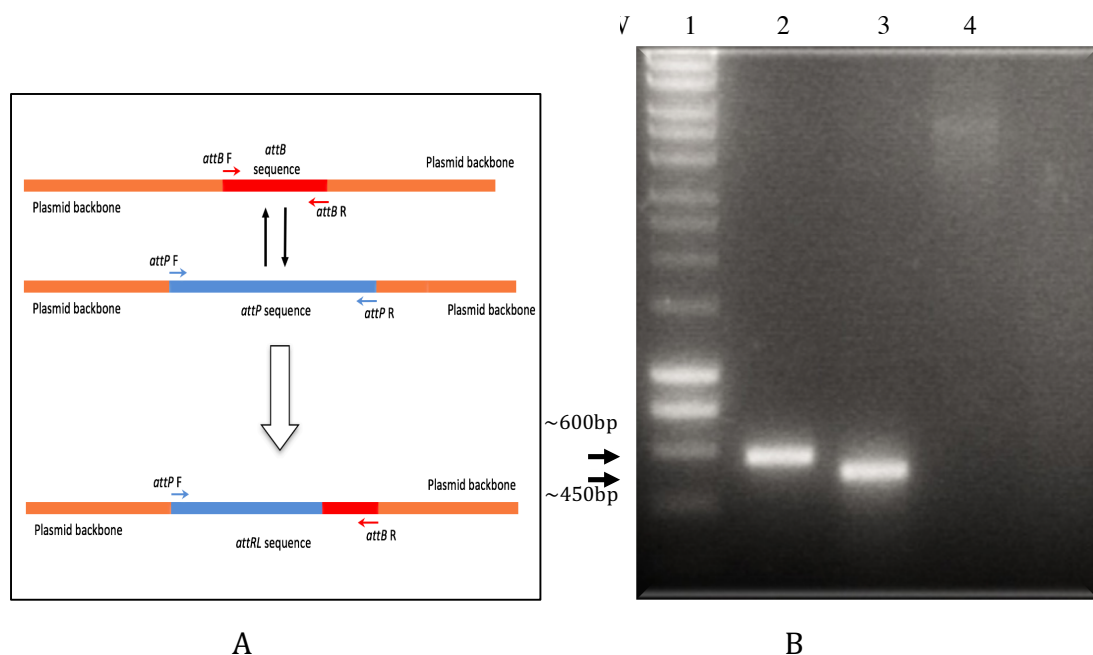


Figure 4.6. PCR detection of the resultant *attRL* hybrid site formation from minimal *attP* sequence detection assays. **A.** Schematic diagram showing the sites of the backbone *attB* F/R primers (Red color) and *attP* F/R primers. The bottom line shows the *attRL* recombination sequence and its F/R primers that used to generate this PCR reaction. **B.** The electrophoresis agarose gel picture Lanes: MW, 1 kb Hyperladder I (Bioline); 1, recombination product of *pattP* 600bp with *pattB* 600 bp; 2, recombination product of *pattP* 400bp with *pattB* 600bp; 3, recombination product of *pattP* 250bp with *pattB* 600 bp; 4, recombination product of *pattP* 200bp with *pattB* 600bp.

4.3.2. The minimal necessary *attB* size

Top 10 competent cells harboring pΦ24_B-attP₁ and pΦ24_B-int were used to determine the minimal necessary *attB* sequence for integrase to drive recombination. These cells were transformed with plasmids harbouring *attB* DNA fragments, each for separate recombination assay (Figs. 4.7).

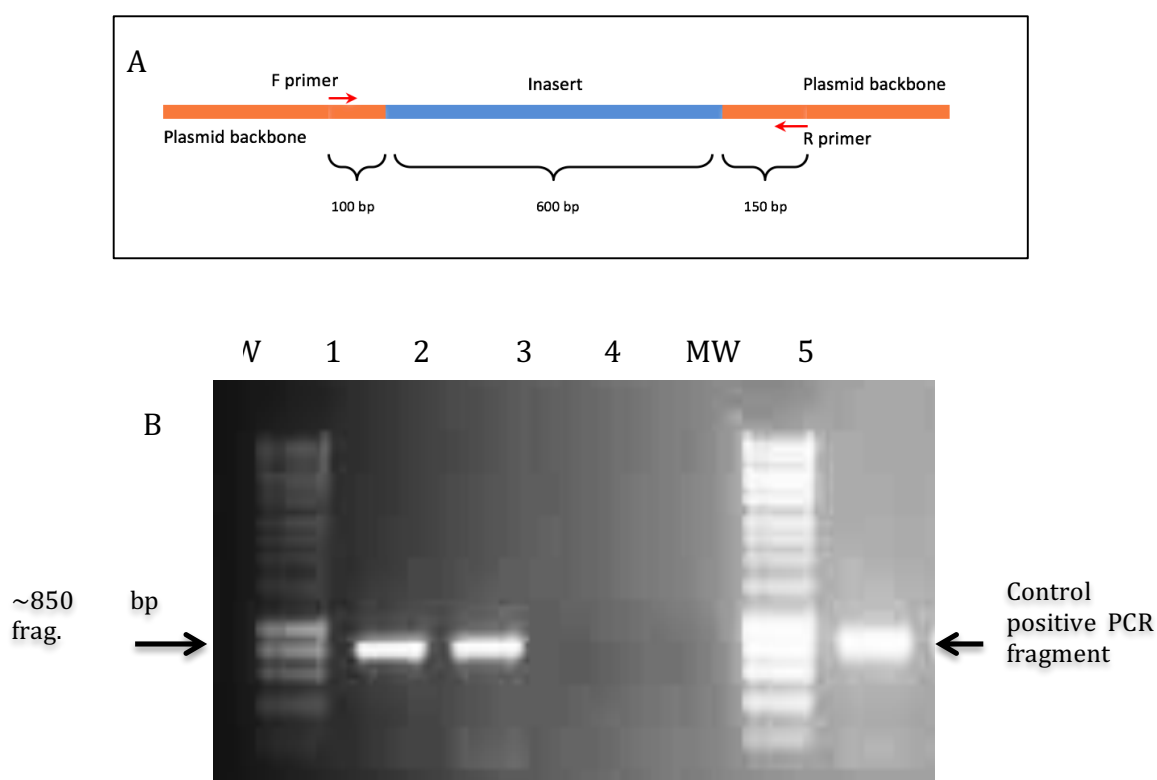


Figure 4.7. PCR results from the *E. coli* Top 10 transformants to confirm the presence of pACYC-Duet-*attB* 600 bp. Top10 cells harboring pΦ24_B-int were transformed with pACYC-Duet-*attP*600. **A.** A schematic diagram showing the sites of the backbone primers (Red color) that used in the PCR reaction, the PCR products were generated with the Forward ACYCDuetUP1 and Reverse T7 Terminator; **B.** The electrophoresis agarose gel picture. MW, 1kb Hyperladder I (Bioline); Lanes 1-2, PCR ~850bp fragments from transformants; Lane 3, negative result colony PCR; Lane 4, negative control; Lane 5, positive control: :PCR for the plasmid with its insert before transformation .

For these assays, the restriction-free cloning method was used. DNA fragments of either 25 bp, 64 bp, 82bp, 93 bp, 110 bp, 125 bp, 200 bp or 300 bp (all carrying the 24 bp core) were cloned, separately, inside the medium copy number pACYC-Duet (Table 2.2.1.). The 600 bp *attB* sequence was cloned in the same plasmid using standard restriction enzyme digests and ligation.

The *in situ* recombination assays were performed in two replicates. The integrations were detected by identifying the presence of large integrated plasmids pΦ24_B-*attRL*. After a plasmids prep of 5 ml assay sample, The outcome was determined by running a PCR reaction using *attRL* F/R primers, on an agarose gel and sending the samples for sequencing. The results showed that the minimal *attB* sequence that Int required to perform integration was 93 bp of DNA (Fig 4.8).

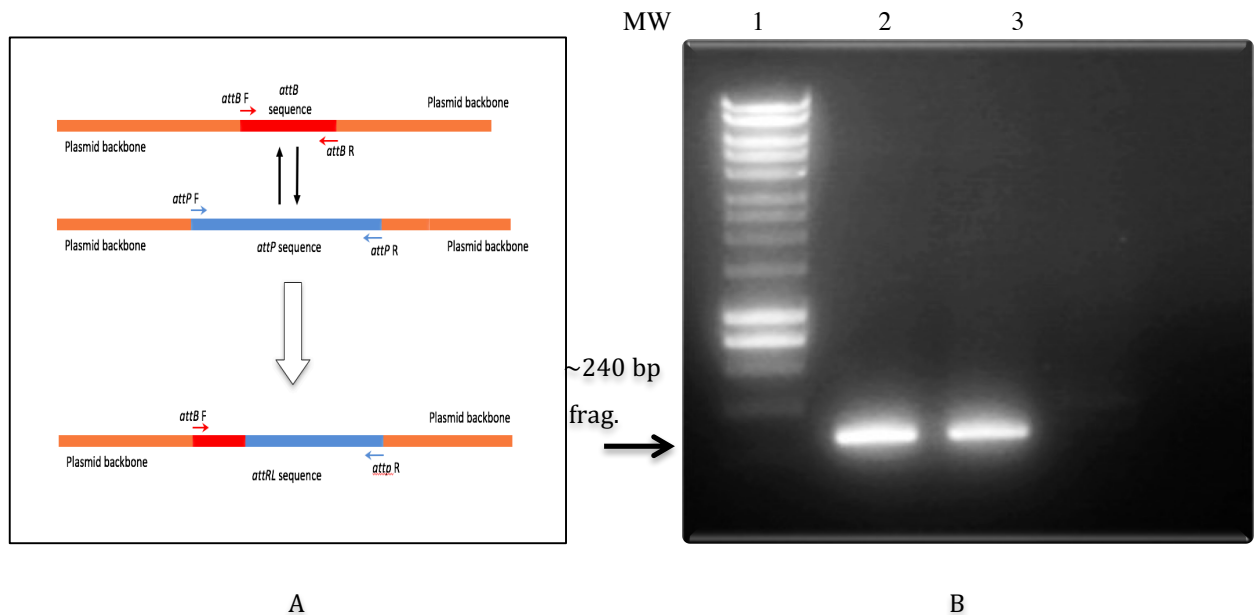


Figure 4.8. PCR detection of the resultant *attRL* sites from minimal *attB* sequence *in situ* integration assays. **A.** Schematic diagram showing the sites of the backbone *attB* F/R primers (Red color) and *attP* F/R primers. The bottom line shows the *attRL* recombination sequence and its F/R primers that used to generate this PCR reaction. Primers that used to check the recombination results are *attB40F/attP200R*. **B.** The electrophoresis agarose gel picture Lanes: MW, 1 kb Hyperladder I (Bioline); Lane 1, recombination product of *pattP* 400bp with *pattB* 100bp; Lane 2, recombination product of *pattP* 400bp with *pattB* 93bp; Lane 3, recombination product of *pattP* 400bp with *pattB* 60bp.

In order to identify whether the sequences that flanking the *attB* site must be present to support integration and whether those flanking sequences must be of equal lengths, several *pattB* plasmid constructs were made and transformed to new Top 10 competent cells already harboring $\Phi 24_B$ -*attP*₁ and $\Phi 24_B$ -*int* for testing in the *in situ* integration assay. The results of these assays indicated that the flanking sequencing comprising the arms of *attB* did not to be of equal size. In fact it was found that the $\Phi 24_B$ -Int needs 49 bp of flanking sequence on the B side

of the core region of the *attB* site and 21 bp flanking sequence on the B' side of the core site in order to drive integration (Fig 4.9).

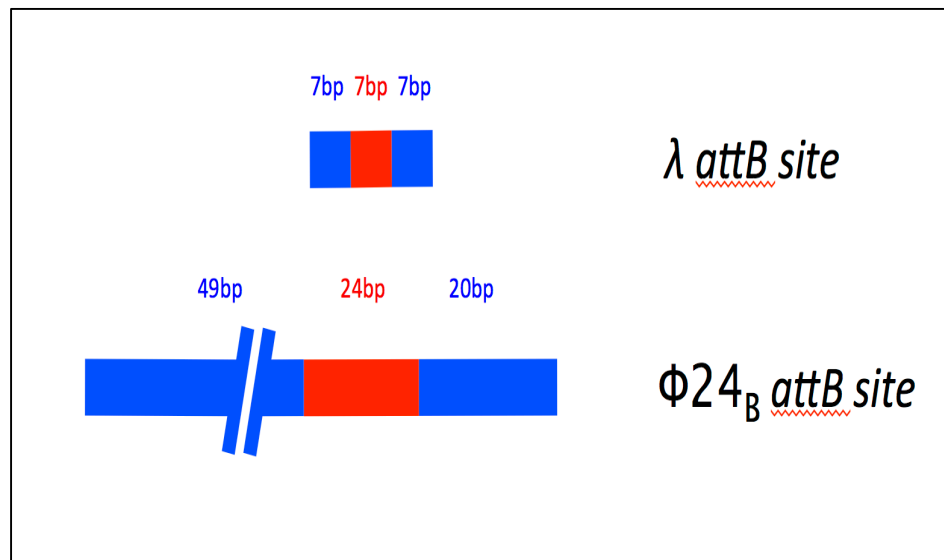


Figure 4.9. Comparison between the essential sequences of the λ *attB* site & $\Phi 24_B$ *attB* sites.

4.3.3 IHF importance determination

To demonstrate the importance of IHF for Int to promote the integration of $\Phi 24_B$ bacteriophage's genome's into *E. coli* genome, the *E. coli* strain JW1702-1 was used as the host cell in the *in situ* recombination assay. This strain is a deletion mutation of the *himA* gene (one of two IHF genes). The assays results showed that a large hybrid integrated DNA product could be detected after arabinose induction of integrase expression, demonstrating that integration could occur in the absence of IHF (Fig 4.10).

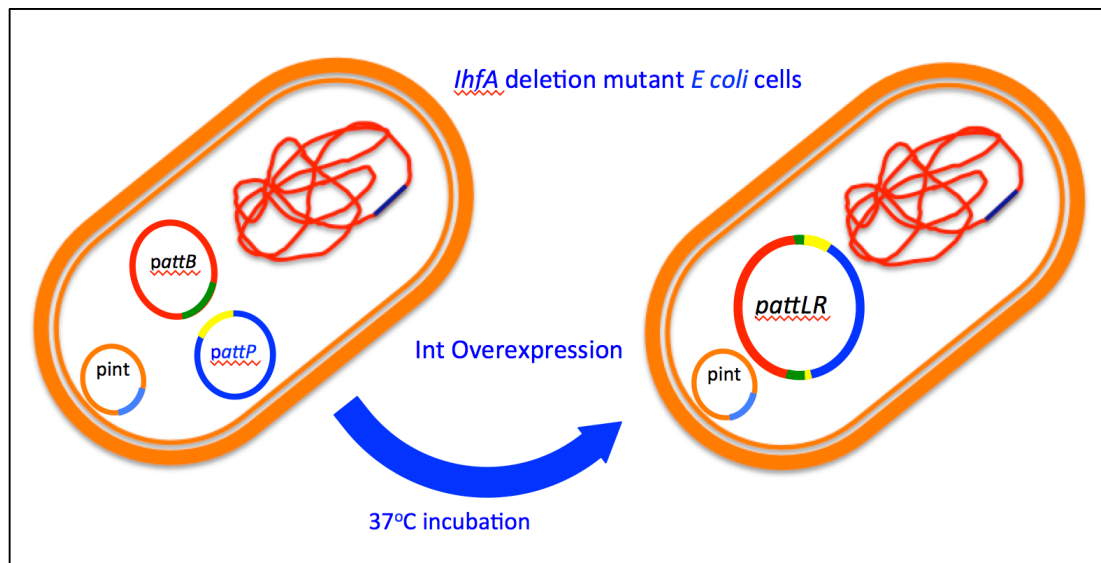


Figure 4.10. *In situ* recombination assay using the *E. coli* strain JW1702-1. The cells harbouring the plasmid mixture composed of $p\Phi 24_B$ -attB, $p\Phi 24_B$ -attP and $p\Phi 24_B$ -int, were incubated until the O.D.₆₀₀ of (0.5-0.6) was reached. At this point the expression of Int protein was induced with arabinose. Int protein, in turn, drives the recombination between $p\Phi 24_B$ -attB and $p\Phi 24_B$ -attP to produce the hybrid recombination product.

4.4. Discussion

As we were not successful in characterising the $\Phi 24_B$ integrase using an *in vitro* recombination assay, we moved on to try a three-plasmid *in situ* recombination assay, which relied on the use of a series of compatible plasmids. This system was used by Fogg (2011), in order to demonstrate the relationship between $\Phi 24_B$ -Int and $\Phi 24_B$ -Xis in several recombination and excision *in situ* assays. *In situ* (or *in vivo*) DNA recombination techniques were also used in many other studies including determining different cellular enzymes that are involved in several types of DNA recombination and to understand many phage genome integrations inside the bacterial genome (Rajeev et al., 2009; McCulloch et al., 1994; Bliska and Cozzarelli, 1987).

Altering our approach to characterise the $\Phi 24_B$ integrase to incorporate the use of an *in situ* assay enabled us to identify the minimal sequence that must be present in *attP* to support integration. We began the *attP* size determination experiments fragments of 600 bp, and constructed several separate plasmids, each plasmid carrying a different sized variant of the *attP* site, but not shorter than 140 bp. Originally, the reason for not examining *attP* sequence lengths shorter than 140 bp was because the minimal *attP* site needed by lambda integrase for integration is 250 bp (Sylwan et al., 2010; Groth and Calos, 2004). The $\Phi 24_B$ bacteriophage belongs to the same family, so we expected that the $\Phi 24_B$ -Int would have similar needs to those of lambda. Moreover, in order to examine whether the P and P' arms flanking the central recombination core need to be equal sizes or not, two *in*

situ assays were run. The first assay was done with a pCDFDuet-*attP* construct that had a P arm of ~140 bp and a P' arm of ~200 bp surrounding the crossover region. The other experiment used a pCDFDuet-*attP* construct with a P arm of ~200 bp and a P' arm of ~140 bp. Neither of these constructs supported recombination. The P' arm in the lambda *attP* is composed of ~80 bp while, the P arm possesses the integrase binding site and is larger at ~140 bp than P'(Goodman et al., 1999). Further work is need to identify all of $\Phi 24_B$ - P, P', H, H', Xis and Fis sites (Fig 3.2), as they are already identified within lambda *attP* site (Numrych et al., 1992).

In addition to identification of the minimal necessary attP site for $\Phi 24_B$ integration, the *in situ* recombination assay enabled the determination of the minimal *attB* sequence requirements. The strategy for determining the minimal *attB* sequences were examined from fragments as small as 25 bp short (essentially limited to the core overall region between *attB* and *attP*). The rational behind using such a short fragment is that $\Phi 24_B$ Int is a site-specific tyrosine recombinase (James et al, 2001). These integrases have complex binding sites, however the *attB* site is usually less complex then the *attP* site. Lambda *attB* site, for example, is only 21 bp with 7 bp serving as the central crossover core sequence that interacts with the *attP* site (Mizuuchi et al., 1981). Nevertheless, according to our results, $\Phi 24_B$ -*attB* site is not that simple. The 25, 64, and 82 bp *attB* sequences failed to support integration. Recombination was however, detected when the *attB* site was 125 bp and the 24 bp crossover region was

flanked by ~50 bp sequences in both the B and B' arms. This sequence size is very large compared to other characterised tyrosine recombinase attB sites (Lorbach et al., 2000, Kolot et al., 1999). The non-integrase tyrosine family members, such as FLP and Cre, need attB sequence ~50 bp, which is still less than what we determined for the $\Phi 24_B$.attB site (Abremski and Hoess, 1984, Andrews et al., 1985).

To identify whether the B and B' arms are equal in their importance in supporting integration, several *in situ* assays were run with pACYCDuet-attB carrying ~50 bp on the B arm side and from 21-40 bp on the B' arm. The large hybrid plasmids, indicated that recombination occurred, were identified in attB site constructs totalling 93 bp and mixed arm sizes with the B arm at 49 bp and the B' arm at 21 bp. This is an unusual result, as most previous reports characterising the requirements for activity of tyrosine-mediated site specific recombinases have identified attB sites possessing B and B' arms of equal size (Lorbach et al., 2000, Kolot et al., 1999).

$\Phi 24_B$ Int is a site-specific tyrosine integrase. All members of the tyrosine family require host encoded DNA binding proteins to perform integration (Groth and Calos, 2004). $\Phi 24_B$ bacteriophage belongs to the same family as lambda phage integrase (Allison, 2007), so we assumed that $\Phi 24_B$ Int would be the same or have requirements similar to lambda Int. According to all that, our hypothesis was that $\Phi 24_B$ Int cannot drive DNA recombination without the presence of the host

encoded, Integration Host Factor (IHF). However, the *in situ* recombination assay performed in *E coli* cells that do not produce IHF proved that DNA recombination can still be driven by $\Phi 24_B$ -Int. This novel finding enables us to rejected the hypothesis of the need for IHF. However, this result does not mean $\Phi 24_B$ -Int can drive recombination without host encoded DNA binding proteins (see. Chapter 3: *in vitro* recombination assay).

CHAPTER 5. IDENTIFICATION OF INTEGRATION SITE PREFERENCES

5.1 Background:

Most temperate *E. coli* phages can integrate in their host's genome via site-specific recombination (Groth and Calos, 2004). They can integrate their genomes in a range of sites within the host genome, however, each phage species has high preference to one specific site (Rajeev et al., 2009). When Lambda phage, for instance, lysogen cells that lack the most preferred attachment site, it can occupy one of the many secondary sites (Rutkai et al., 2003). These secondary sites have a degree of homology with the primary site in their core overlap region, flanking sequence or both (Richet et al., 1988). However, the affinity to these sites is not as strong as that for the primary attachment site, and therefore their utilization frequency is less than that for the primary attachment site. Integration specificity depends on protein-DNA interactions between the integrase molecule, the phage DNA attachment site (*attP*) and the bacterial DNA attachment site (*attB*), as well as the DNA/DNA interaction between the corresponding overlap regions of *attP* and *attB* (Weisberg et al., 1983).

When Lambda phage directs its integration into the host's genome to become a prophage it produces a repressor (CI), which is controlled by a self-regulated promoter, P_{RM} . CI's role is to inhibit the expression of all other phage genes in the lysogen and provide immunity to superinfection by other Lambda phages (Lisio and Weissbach, 1965). Therefore a double lysogen of Lambda phage have only

been found in *E. coli* host cells that carried a Lambda with immunity mutant prophages (Freifelder and Kirschner, 1971).

The bacteriophage, $\Phi 24_B$, is a lambdoid phage that gives the enterohaemorrhagic *E. coli* (EHEC) their ability to produce Shiga toxin (Stx) (Paton and Paton, 1998). The lambdoid CI-controlled immunity to superinfection does not necessarily function the same in these phages. Isogenic, selectable, non-toxigenic $\Phi 24_B$ mutants were used to produce a multiply infected host cell (Allison et al., 2003). Derivatives of $\Phi 24_B$ were found to be able to sequentially infect a host cell, each integrating at a distinct site (Fogg et al., 2007). The rate of occurrence of this superinfection increased in frequency with each successive superinfection event. The harboring of more than one isogenic prophage simultaneously within a single bacterial host cell has implications for enhanced pathogenicity as a result of the presence of multiple copies of the Stx genes in a single bacterial cell (Fogg et al., 2010; Fogg et al., 2012; Botstein, 1980). Furthermore, it makes the host cell more sensitive to an induction stimulus than lysogens harboring a single prophage, suggesting that maintenance of a stable prophage is less likely when a superinfection is present (Fogg et al., 2012; Botstein, 1980).

At least four isogenic prophage genomes were identified following an experiment involving the sequential infection of a $\Phi 24_B$ lysogen, with each prophage integrated at a unique site (Fogg et al., 2007). The *attB*₁ is just upstream of the integrase gene from the cryptic prophage *intS*, so it can be called *attB*_{intS} site. The

site *attB₂* lies between the genes *yfaL* and *nrdA*, encoding an adhesion and ribonucleoside-diphosphate reductase, respectively, therefore, it can be named *attB_{yfnr}* site. The site *attB₃* lies within the *yfbO* gene encoding a hypothetical protein, so it can be named *attB_{yfbO}* site. The site *attB₄* is within *sgcA* gene encoding a putative phosphotransferase enzyme IIA component, so it can be named *attB_{sgcA}* site. With the exception of the *in* integration site, which is favoured, there seems to be no preference (Fogg et al., 2007) (fig.5.1).

Since it was determined that $\Phi 24_B$ -phage can drive multiple integration events (Allison *et el*, 2003), within at least four sites, the four sites were sequenced (Fogg *et el*, 2007). The alignment of the $\Phi 24_B$ genome sequence (*attP*), the MC1061 genome sequence (*attB*) and the integrated prophage (within lysogen genome) sequence (*attL*) clearly showed that the primary bacterial attachment site has 100% homology to phage attachment sites. On the other hand, the alignment among the remaining bacterial attachment sites showed that the similarity among them is about 50% (Fig.5.1). This is definitely likely to explain the answer for why *in* attachment site always gives lysogen first.

| Site | O sequences | Near loci | Genome location (W3110 ref. seq) |
|----------------------------|--------------------------|------------------------|--------------------------------------|
| <i>attB_{intS}</i> | TTATATCCATTTAACTAAGAGGAC | <i>intS</i> | 2,471,991 |
| <i>attB_{yfbO}</i> | GCTTATCTGTTTATATAAATGTTT | <i>yfbO</i> | 2,393,251 |
| <i>attB_{yfnr}</i> | CTATATCAATTTATCTAACCTATT | <i>yfaL & nrdA</i> | 2,348,839 & 2,349,535 |
| <i>attB_{sgcA}</i> | CTGTACCCATTTAATATCGTTAAT | <i>sgcA</i> | 4,532,229 |

Figure 5.1: The alignment of four distinct bacterial integration sites in same genome.

Site *attB_{intS}* is the primary attachment site, which has 100% homology with phage attachment site. Site *attB_{yfbO}* has 62.5% to the primary attachment site and to phage attachment site. Site II has 62.5% to the primary attachment site and to phage attachment site. Site *attB_{yfnr}* has 50% to the primary attachment site and to phage attachment site. And finally Site *attB_{sgcA}* has 54% to the primary attachment site. DNA nucleotides that are highlighted with red colour are present in all of the four sequences. Yellow color for that present in three of them, while sequences that exist in two of them are highlighted with green color.

It has been reported that *E. coli* cells can be simultaneously the host of multiple compatible plasmids with a high DNA concentration (see 4.3. *In situ* assay, Chapter 4). This co-transformation is found due to the description of high efficiency *E. coli* electroporation. Multiple vector transformants have been described in previous studies (Dower et al., 1988). The ability of simultaneous transformation of compatible plasmids played an important role during the bacterial constructions.

Cells with multiple plasmids have been used in previous studies for a variety of purposes, for example: to determine differences in the expression of mutant genes

compared to wild type genes (Peixoto et al., 2004). The outcome of such studies can be determined by using fluorescence-based quantitative Real-Time PCR (qPCR) to measure the number of copies of RNA transcripts of both the mutant and the wild type genes (Bustin, 2000). Other studies have used this qPCR to address the presence of some genes even without expression, utilizing the high specificity of this quantification technique, using primers that targeting their DNA sequences directly. qPCR is particularly suitable here, as even very small DNA copy number can be amplified and identified (Jonker et al., 2008; Rooks et al., 2010). Real-Time PCR is a molecular technique based on the polymerase chain reaction (PCR), however it determines the targeted DNA amplification during the PCR, and not at its end as in conventional PCR. To obtain this feature, syber green, which interacts with dsDNA and marked with a fluorophore (a fluorescent chemical compound), is added to the reaction mixture in a thermal cycler contains sensors for measuring the fluorescence of the fluorophore after it has been created at required wavelength. This allows the generation rate of PCR amplification to be measured for one or more specific products at each PCR cycle. The generated data can be analysed by computer software (Excel Microsoft program) to calculate the mRNA or the DNA copy number in several samples (Bustin, 2000).

qPCR is a powerful tool, with its high capacity of determine and measure small amounts of the nucleic acid in different samples from variety resources. Its specificity along with its speed combination have made it criterion for nucleic acid quantification. In addition to the use of qPCR as a research instrument, several medical diagnostic applications have been performed depending on the

quantification of DNA, for example: risk estimation of cancer recurrence, the identification of gene dosage, and forensic applications (Huggett et al., 2013).

In this study we have used qPCR to determine the outcome of the *in situ* recombination assay, which is performed inside *E. coli* cells that are transformed with more than four compatible plasmids. qPCR requires a reference gene to make comparisons with, and as the phage attachment site was going to be present in all of recombination products, the marker gene of the plasmid CDFDuet that harboring *attP* site in all experiments was used as a reference gene for our qPCR assay.

5.1.1 The Aim

- To identify the recombination events within each bacterial attachment site compared to the primary bacterial attachment site.
- To identify the frequency and the order of preference for each *attB* site over time.

5.2. Specific Methods

5.2.1. Bacterial strains, plasmids and media

The bacterial strains and plasmids used in this study are listed in Table 2.2.1 (Chapter 2). One Short TOP10 *E. coli* cells were used, as a host, for all *in situ* recombination assays. The genome of *E. coli* K-12 strain, MC1061, was used as a template for the amplification of all four bacterial attachment sites (*attBs*). Each *attB* site was cloned into a plasmid with a unique replicon (Table 2.1.1). The phage attachment site (*attP*) was amplified from the genome of the MC1060/Φ24_B lysogen and cloned in its appropriate plasmid (Table 2.1.1). MC1060Φ24_B refers to MC1061 when it carries the Φ24_B prophage, while naïve MC1061 refers to the strain when it has not been infected by Φ24_B. The routinely used culture media were Luria-Bertain broth, Luria-bertani agar and SOC Growth Media. The following antibiotics were used where appropriate: chloramphenicol (50 µg ml⁻¹), tetracycline (10 µg ml⁻¹), spectinomycin (100 µg ml⁻¹), ampicillin (100 µg ml⁻¹), kanamycin (50 µg ml⁻¹), and gentamicin (10 µg ml⁻¹).

5.2.2. PCR cloning

The overlap extension PCR cloning was performed as mentioned previously (See 4.2.1.2 Restriction-free cloning method, Chapter 4)(Bryksin and Matsumura, 2010). Specific primers that have insert sequence at the 3' and plasmid sequence at the 5' end were used to amplify the insert sequence first (Table 2.5.2. Chapter 2). The second step of this process is to use these inserts as mega-primers to amplify

the plasmid backbone in a secondary PCR reaction. The Highly Fidelity Q5 high fidelity polymerase was used in both first and second PCR reactions. PCR conditions for both of the reactions initial denaturation step at 98 °C for 30 sec., followed by 30 cycles of denaturation at the same temperature for 10 sec., primer annealing time was 15 sec at the specific primer's temperatures. The PCR extension took place at 72 °C. It's time was 30 sec. per kb product. The original plasmid templates in the PCR products, were digested using *DpnI* digestion enzyme (Biolabs). The final products were transformed to new One Short TOP10 *E. coli* competent cells.

5.2.3. *In situ* recombination assay.

All *in situ* recombination assays were performed inside One Short TOP10 *E. coli* cells. After the transformation of each plasmid, new competent cells were created to be ready for receiving the next plasmid until all of compatible plasmids with different markers were all in one *E. coli* host cell. The plasmids are five different compatible plasmid harboring four different *attB* sites and *attP* site along with plasmid harboring *int* gene (Fig.5.2.).

An 100 ml amount of LB broth was inoculated with 1 ml of an overnight culture of cells harboring the compatible plasmids. This was incubated at 37 °C with shaking at 200 r.p.m to the OD₆₀₀ 0.5-0.8. Integrase protein Int was overexpressed by adding L-arabinose to a final concentration of 0.02%. The cultures were incubated over various times, as indicated. From each assay flask, 5 ml samples were taken

for plasmid extraction. All assays were performed along with a negative control where the plasmid bearing cells were not exposed to arabinose to repress the expression of the integrase protein.

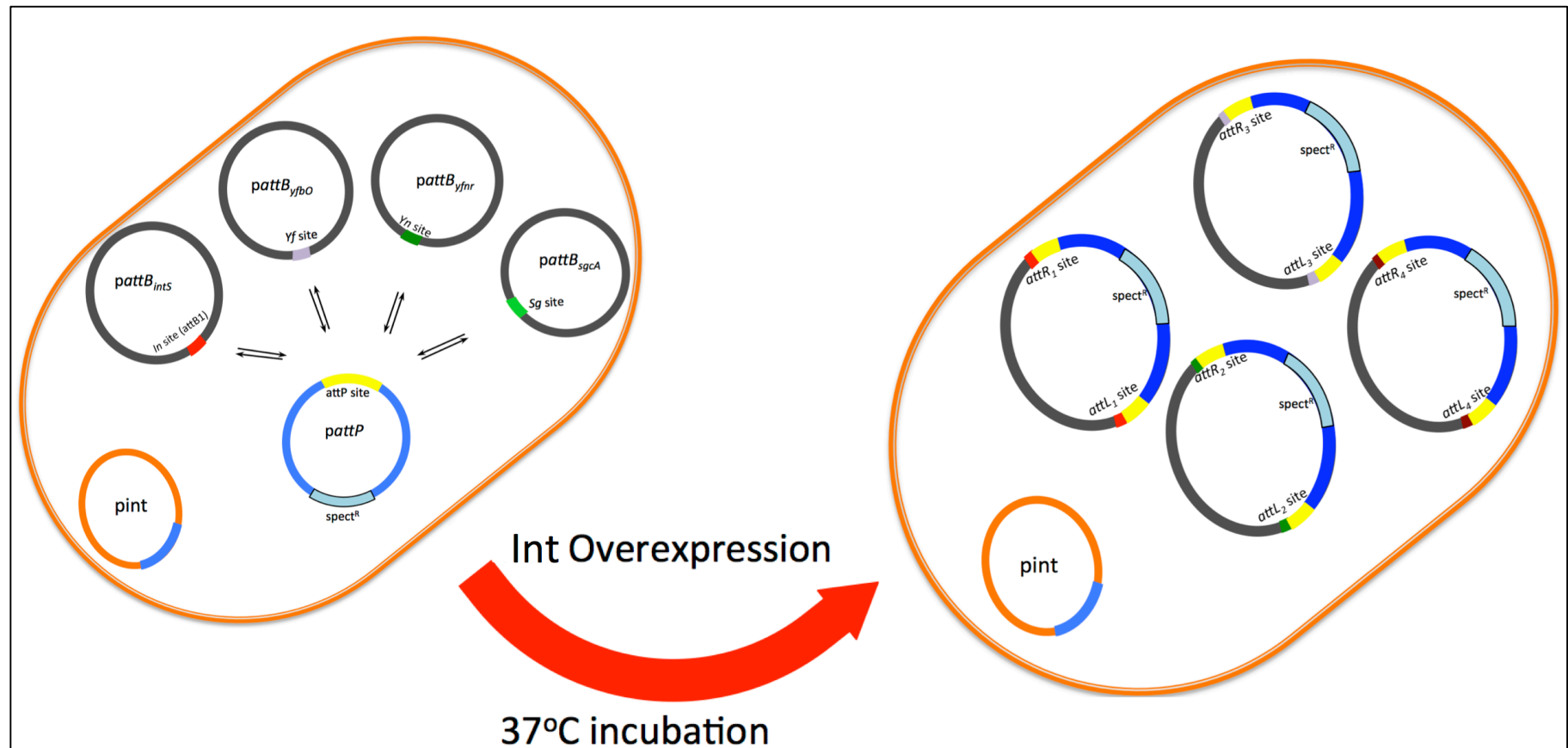


Figure 5.2. The substrates and the final recombination products of *in situ* recombination assay. After inducing of Int protein overexpression from its pBAD plasmid (Orange color), DNA recombination takes place between *attP* site (yellow color) in its plasmid (Blue color) and each of *attB* sites in there plasmids (Black color). Each recombination event takes place between each *attB* site and *attP* site gives specific product differs from other *attB* site recombination products. Depending on that we could have specific primers for each hybrid recombination product to be used for finding out the difference between the recombinations events in qPCR technique. The reference gene that used in the qPCR assays was the spectinomycin resistant gene (*aadA2*) (Light blue color), as it is present within all of the recombination products.

5.2.4. Comparative qPCR assay.

In situ recombination assays cultures were harvested for plasmid mini preps (Bioline), according to the manufacturer's specifications. The reference gene that was used in the qPCR assays was the spectinomycin resistant gene (*aadA2*), as it is present within all of the recombination products (Fig.5.2.). Specific primers were used for all of the recombination's products for each qPCR reaction in SensiFAST SYBR qPCR reaction mixture (Bioline). All experiments were performed to comply with the standard cycling conditions. The experiments' annealing temperature was 53 °C and the amplicon size was ~200 bp for all of recombination products. The qPCR assays were performed using the Step One Plus Real-Time PCR Cycler (Applied Biosystems, Waltham, USA)(Fig.5.2). The efficiency of the qPCR primers was confirmed by standard curve analysis for each pair of primers (Copy number range = 3×10^4 - 3×10^8). The qPCR outcomes were confirmed by running them on a agarose gel. The DNA band of interest was cut, purified and sent for sequencing. The typical thermo cycling qPCR program started with 20 sec. initial denaturation step at 95 °C. The cycling stage was composed of second denaturation step at 95 °C for 2 sec. followed by 10 sec. annealing step, and extension step at 72 °C for 20 sec. A melting curve analysis was performed for each amplification reaction, with a temperature gradient of 1 °C from 53 °C to 95 °C. All qPCR assays were performed with a negative control which has exactly same conditions but without template. The qPCR resulted data were analyzed using Delta Delta C_T method to calculate the quantitative results of each recombination event and compare it with

In site recombination results. The mathematical equation that was used to find out the Delta Delta C_T for each recombination event was:

$$2^{-\Delta C_T} = (C_T \text{ of reference Gene} - C_T \text{ of } In \text{ site recombination product}) - (C_T \text{ of reference Gene} - C_T \text{ of tested site recombination product})$$

C_T is the PCR cycle at which the fluorescent signal of the reporter dye crosses an arbitrarily placed threshold. The t-test analysis was used to find out the significance of the qPCR results

5.3. Results

In order to have all $\Phi 24_B$ bacterial attachment sites cloned in separate compatible plasmids, all of the *attB* sites *attB_{ints}*, *attB_{yfnr}*, *attB_{yfbO}* and *attB_{sgcA}* were cloned inside pCYCDuet, pRSFDuet, pETDuet and pCOLADuet plasmids, respectively, to create p $\Phi 24_B$ -*attB₁* (*pattB_{ints}*), p $\Phi 24_B$ -*attB₂* (*pattB_{yfbO}*), p $\Phi 24_B$ -*attB₃* (*pattB_{yfnr}*), p $\Phi 24_B$ -*attB₄* (*pattB_{sgcA}*) and p $\Phi 24_B$ -*attP* (*pattP*) (Table 2.2.1)(Fig.5.2.) .

Unfortunately, both *pattB_{yfnr}* plasmid and *pattB_{sgcA}* plasmids possessed the kanamycin resistant gene (*kan^{R2}*) as a selectable marker. To enable selection for both plasmids simultaneously, the marker in *pattB_{sgcA}* was replaced by the tetracycline resistant gene (*tet^R*). The tetracycline resistance gene (*tet^R*) was amplified from the plasmid pBR322 and cloned into pCOLADuet-Sg. Furthermore, the ampicillin resistant gene (*bla*) in *pattB_{yfbO}* was replaced with the gentamicin resistant gene (*aacC1*) as plasmid marker. The phage attachment site (*attP*) was cloned into pCDFDuet that carries the spectinomycin resistant gene (*aadA2*) which served as the indicator gene for the qPCR assays, being found in the plasmids bearing *attP* and the recombined plasmids (Table 2.2.1). All plasmids were transformed into one bacterial host.

5.3.1. Relative quantitative recombination events

To determine the quantitative recombination events of all secondary $\Phi 24_B$ bacterial attachment sites, the number of recombination events occurring within the Top10 cells was compared to the number of recombination events within the

primary attachment site. Two variations of *in situ* assays were performed in two replicates for each: first; *pattB_{intS}*, *pattB_{yfnr}* and *pattB_{yfb0}* were transformed into OneShotTOP10 competent cells along with both *pattP* and pΦ24_B-*int* (plasmid that harboring integrase gene); second; cells harboring *pattB_{intS}*, *pattB_{yfnr}*, *pattB_{yfb0}* and *pattB_{sgcA}* as well as *pattP* and pΦ24_B-*int* were used. Cell cultures harboring these plasmid mixtures took longer to grow (6-8 h) before reaching an OD₆₀₀=0.5-0.8. After reaching this optical density the integrase gene was induced with arabinose to drive its expression, enabling recombination. After 14 hours, plasmids were recovered from the cells and recombination events between *attP* and the *attB* sites were detected by qPCR (Fig. 5.3).

The qPCR data showed there were a significant quantitative difference between the recombination events in the primary attachment site (*attB_{intS}*) (*attRL₁*) and in other attachment sites (P-value = 0.00) (Fig. 5.3). The formation of *attRL₂* and *attRL₃* were low compared to the formation of the *attRL₁* site, and there was no significant difference between their relative quantities. The fourth-bacterial attachment site recombination *attRL₄*, in the experiment that has all of attachment sites plasmids in one bacterial host cell, showed the same relationship to the primary attachment site, as there was a significant lower number of *attRL₄* compared to *attRL₁*. Furthermore, it showed no significant formation Differences were detected across *attRL₄*, *attRL₂* and *attRL₃*.

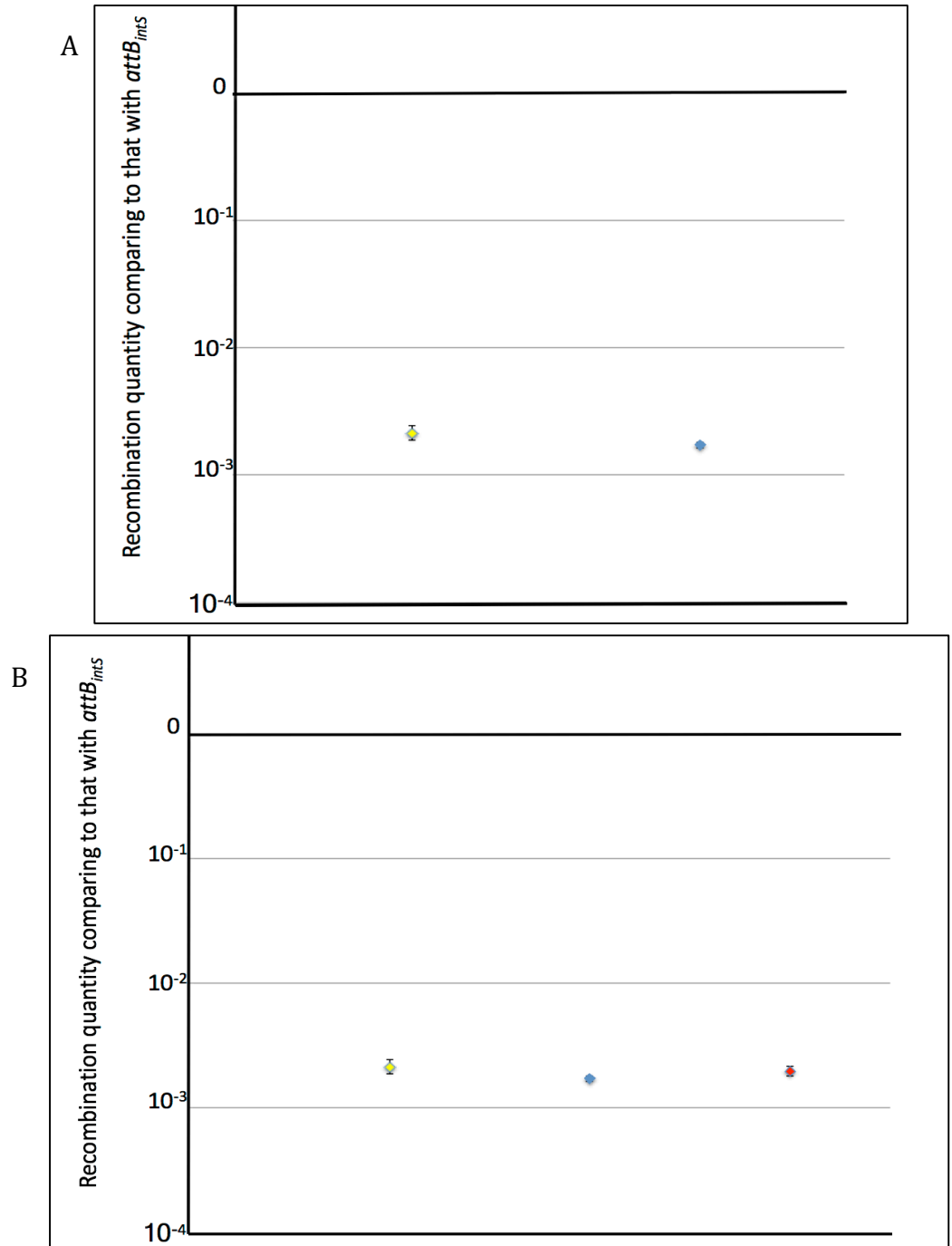


Figure 5.3. The relative quantitative recombination events in the secondary sites mediated by $\Phi 24_B$ Int measured by qPCR as compared to the use of $attB_{intS}$. Expression of integrase was driven by an arabinose inducible promoter at 37 °C after the culture reached mid-exponential growth, 14 hours later, plasmids were harvested from the cells and recombination was determined by qPCR.

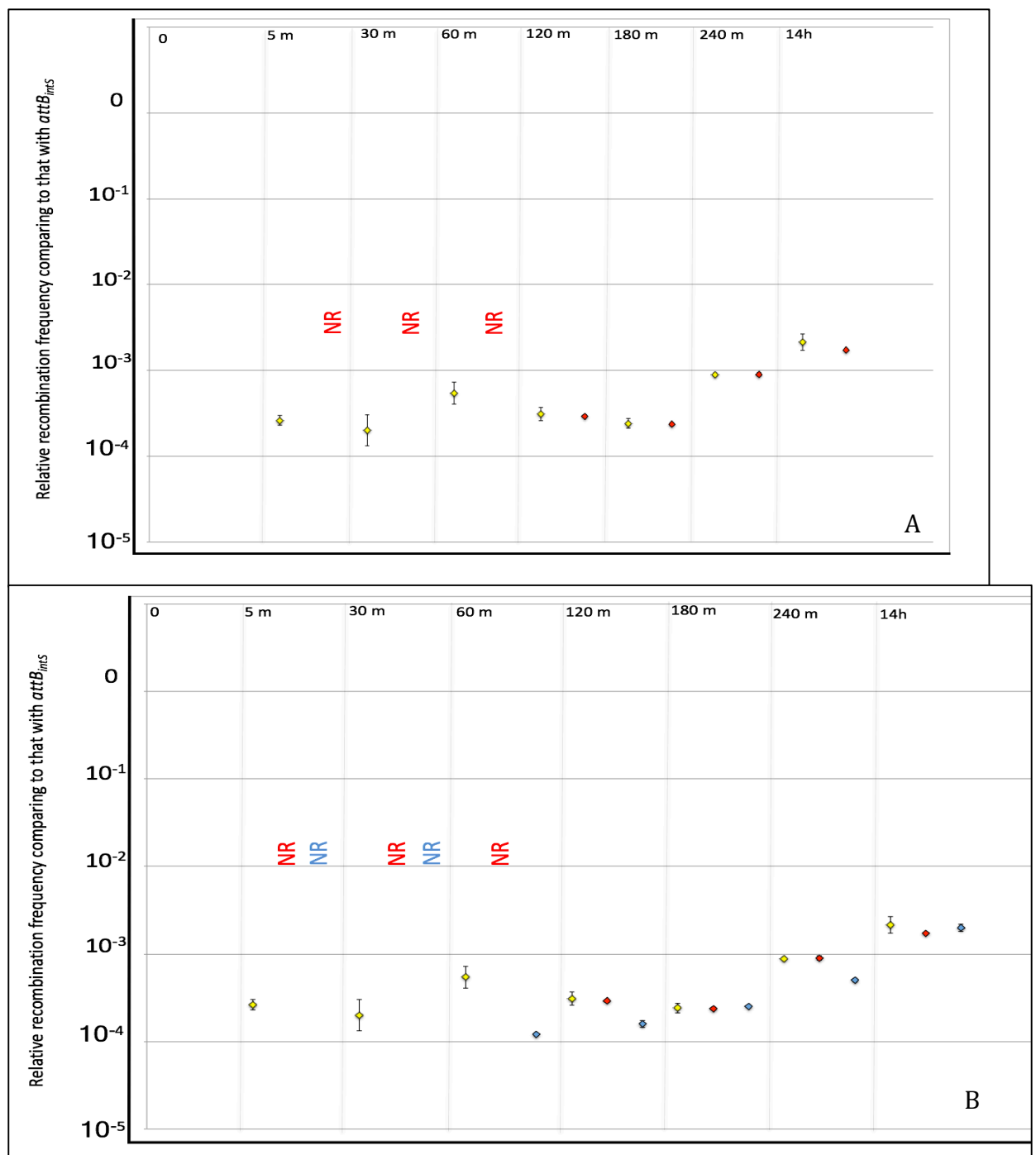
A. The qPCR results of the recombination assay of cells harboring $pattB_{intS}$ (labeled by number 1), $pattB_{yfnr}$ (yellow color) and $pattB_{yfb0}$ (blue color).

B. The qPCR results of the recombination assay of cells harboring $pattB_{intS}$, $pattB_{yfnr}$, $pattB_{yfb0}$ and $pattB_{sgA}$ (red color).

The *in situ* recombination events of $attP$ with $attB_{yfnr}$, $attB_{yfb0}$ or $attB_{sgA}$ were significantly less frequent than the recombination with $attB_{intS}$ (p-value= 0.00). The frequency of $attP$ recombination with in the secondary sites was similar, with no significant differences detected.

5.3.2. Recombination site preference of *attP* site

The presence of $\Phi 24_B$ bacterial attachment sites in different compatible plasmids, having different plasmid markers, in one *E. coli* bacterial host cell enabled us to find the difference among these recombination sites not only in their integration frequency but also in their recombination preferences. The preference of each bacterial attachment site's integration could be determined by performing a time course experiment. Time course could be demonstrated by taking 5 ml samples from the *in situ* recombination bacterial culture after 5 m, 30 m, 60 m, 120 m, 180 m, and 240 m, after integrase protein was induced. The culture was then left for a further 10 h before it was harvested for final sample. These samples were taken for plasmid mini prep and qPCR experiments. The time course qPCR of the *in situ* recombination assay that has *pattB_{intS}*, *pattB_{yfnr}* and *pattB_{yfbO}*, showed that the first recombination events took place with *attB_{intS}* and *attB_{yfnr}* after 5 m only. While, for integrase to drive the *attRL₃* recombination, 120 m was needed. Similar results were seen in the time course qPCR of the *in situ* recombination assay, that has *pattB_{intS}*, *pattB_{yfnr}*, *pattB_{yfbO}* and *pattB_{sgcA}*, for all *attRL₁*, *attRL₂* and *attRL₃*. The experiment results showed that *attRL₄* took place within 60 m, before *attRL₃* (Fig.5.4).



5.4. Discussion

Temperate phages that enhance the pathogenicity of their hosts are known as converting phages (Saunders et al., 2001). These phages play an important role in the dissemination of virulence and other genetic determinants that increase the genetic diversity across members of bacterial populations. These phages can even transform nonpathogenic bacteria into causative agents of several diseases for plant, animal or human hosts (Allison, 2007).

Stx-phages enhance the evolution and virulence of its *E. coli* pathogens (Brüssow et al., 2004; O'Brien et al., 1984). The pathogenicity enhancement is massively increased by multiple Stx prophages in a single bacterial host. Moreover, multiple lysogens are playing important role in the dissemination of phage-encoded genes among the bacterial populations (Allison *et al*, 2003)(see 1.3 Temperate Phage and HGT, chapter 1). The temperate phage, $\Phi 24_B$, confers upon its host the ability to produce Shiga toxin. It has been suggested that the presence of more than one copy of Shiga toxin encoding genes (*stxAB*), due to the multiple lysogens, might be the key virulence factor behind the capability of Shiga-toxigenic *E. coli* of causing a variety of clinical symptoms and sequelae ranging from mild diarrhea through to life threatening conditions such as hemolytic uremic syndrome and thrombotic thrombocytopenic purpura (Fogg *et al*, 2007).

It has been reported that integrase cannot drive integration in the secondary attachment site without the absence of the primary site or the primary site being occupied by a resident prophage (Rutkai et al., 2003). In order to confirm or reject

this hypothesis, we cloned all the different bacterial attachment sites, each in separate compatible plasmids, and put them all into in one host cell along with two other plasmids carrying *attP* and the integrase gene, respectively. After running the *in situ* recombination assay for these host cells, total plasmid DNA was harvested and qPCR was used to confirm the ability of the integrase to catalyze recombination of *attP* with all of the bacterial attachment sites, concurrently.

$\Phi 24_B$ integrase expression is not controlled by the phage repressor, which is different from Lambda repressor that is capable of completely shutting down lambda integrase expression by turning off P_L . The integrase from $\Phi 24_B$ is expressed continuously as the integrase genes and its promoter lie in the opposite direction to P_L . Therefore each successive prophage integration results in increased expression levels of the $\Phi 24_B$ integrase enzyme. This continuous production of integrase protein leads to its intracellular accumulation in the cytoplasm. Even though endogenous repressor is being made from the resident prophage, which may block the expression of genes associated with newly infecting phages, the transacting nature of the integrase enzyme will direct the integration of the infecting phage into an available *attB* site. The increasing levels of *trans* acting integrase in multiple lysogens also explains the increase in lysogen forming frequency of superinfected lysogens compared to naïve host infection (Fogg et al., 2011). Induction can still occur for $\Phi 24_B$ because the Xis is controlled

by PL, and will be expressed upon the RecA mediated autocatalysis of $\Phi 24_B$ repressor (Fogg et al., 2011).

We determined the quantitative recombination events of each bacterial secondary integration sites $attB_{yfnr}$, $attB_{yfbO}$ and $attB_{sgcA}$ and compare it to the primary attachment site's recombination formation $attB_{intS}$, the first site that is recognized by the phage integrase protein. It is clear from our data that the quantitative formation of double and multiple plasmids recombination are significantly less than that with the attachment site $attB_{intS}$. Nevertheless, the integration formation at supplementary $attB$ sites were comparable to each other after 14h, as there was no significant difference among them.

The presence of four $\Phi 24_B$ bacterial sites in plasmids can utilize recombination with other plasmids having phage attachment site in one bacterial cells can help as to found out the preference of each supplementary recombination. It has been reported that the primary attachment site $attB_{intS}$ has 24 bp sequence of 100% identity with $\Phi 24_B$ $attP$, so without any doubt, this site is the most preferred as it is recognized very fast by the integrase protein. The alignment supplementary sites sequences to the primary attachment site, showed that the similarity percentage to the $attB_{yfnr}$ site is 62.4%, while site $attB_{sgcA}$ was 54% similar to site $attB_{intS}$, and 50% similarity for site $attB_{yfbO}$ (Fogg et al., 2011). These data can be the reason behind the existence of $attRL_1$ and $attRL_2$ within just 5m. And this is giving an explanation why $attRL_4$ always coming before $attRL_3$, as the similarity of $attB_{sgcA}$

site is slightly higher than that of $attB_{yfbO}$ so it can be faster recognized by the integrase. It is also clear from our data that the frequency of attRL₂ recombination was clearly increased from its starting point at 5 minutes to the final time point (14h) of the experiment. This confirms what has been mentioned before about frequency increasing of lysogen formation during superinfection compared to infection of naïve cells (Fogg et al., 2011).

CHAPTER 6. GENERAL DISCUSSION AND FUTURE WORK

6.1. General Discussion

Temperate phages are bacteriophages that have the ability to infect bacterial cells and replicate in one of two pathways: the lytic replication cycle, where the phage uses the bacterial host's machinery to produce new phage particles and that finally leads to the demise of the bacterial host cell; or the lysogenic replication cycle, where the phage can direct its integration into its host cell's genome to become a prophage. In the lysogenic pathway, the prophage can be passed along to the next generation of cells like any other genetic loci. Prophages can, upon some inducing signal, also be reactivated into the lytic cycle. It has been stated that temperate phages have the ability to introduce additional genetic material to their bacterial host (Saunders et al., 2001). These genetic materials, according to several genome sequencing studies, have the potential to encode virulence factors or virulence-associated traits that can alter and enhance the pathogenic profile of the bacterial host (Creuzburg et al., 2005). The phage $\Phi 24_B$ is a temperate phage belonging to the lambdoid family. This phage is a converting phage contributing to the pathogenic profile of its *E. coli* bacterial lysogen, by enabling the lysogenic host to produce Shiga toxin (Allison, 2007).

For the incoming phage to become a prophage, the phage encoded integrase enzyme must be produced to direct prophage integration into the bacterial chromosome and repressor protein must be produced to prevent the expression of the lytic cycle-associated genes (Serra-Moreno et al., 2008).

Lambda Int requires host encoded IHF to promote DNA recombination between *attB* and *attP* sites (Craig and Nash, 1984). We started our experiments with minimal lambda Int requirements for use in our *in vitro* recombination assay development, as $\Phi 24_B$ phage belongs to the lambdoid family of phages. We considered that the $\Phi 24_B$ Int would need similar requirements and factors to drive integration. Before we performed the *in vitro* recombination assay, we had to produce all the needed components. These components were the $\Phi 24_B$ *attP* and $\Phi 24_B$ *attB* sequences, purified $\Phi 24_B$ integrase, and purified *E. coli*-encoded integration host factor. We focused on optimizing the IHF purification, as we were aimed to have a highly-purified IHF protein at a high concentration before initiating a number of *in vitro* assays. IHF is a small heterodimeric protein that assists in a variety of DNA transactions via site-specific binding. A common mechanism underlying the diverse array of processes affected by IHF DNA bending (Nash, 1990).

We aimed in this study to fully understand how $\Phi 24_B$ Int mediates integration and to define the required factors needed for this enzyme to function, which could make it applicable to several biotechnological applications and genetic engineering.

Previous studies have demonstrated that experimental parameters required for various site specific recombinases for *in vitro* recombination assays are not

always the same. Most serine recombinases can drive *in vitro* recombination events without any host encoded accessory proteins (Fogg et al., 2014). $\Phi 24_B$ Int is a site-specific recombinase belonging to the tyrosine family of integrases. With the exception of Cre, encoded by bacteriophage P1, most members of the tyrosine family of recombinases require host encoded accessory proteins to mediate the integration (Groth and Calos, 2004). We considered that IHF would be a necessary factor for $\Phi 24_B$ integrase mediated integration. However, we never detected recombination events in our *in vitro* assay without the addition of a crude *E. coli* cell extract. It was difficult to address why this was the case because our cell extract was made from a culture that overexpressed IHF. This over expression strain was actually the same strain that produced IHF tagged with a poly histine tag for affinity purification. Therefore, further work was required to find out whether our purified IHF protein was biologically active or not, or whether some other unknown accessory protein needed by $\Phi 24_B$ Int to drive recombination. Moreover, recombination was only detected in the *in vitro* assay when *pattB* was linearized, a phenomenon reported by Frumerie *et al.*, (2008), who demonstrated that the use of linear *attB* DNA fragments with supercoiled *attP* plasmid and IHF was necessary to facilitate recombination by the P2 integrase enzyme *in vitro*. In the end, this assay was cumbersome to perform and resulted in only a limited amount of interpretable data.

There are many advantages to the use of an *in vitro* assay to characterise site specific DNA recombination. An *in vitro* assay enables fine control of the amounts

of the DNA, protein and enzymes used in the assay, and subsequently their required concentration and ratios to each other can be determined. The constituents, cofactors and pH of the reaction buffer can be controlled and optimized in order to provide the appropriate condition for the integrase to function. These advantages would enable very precise characterisation of the conditions under which the integrase functions. An *in vitro* assay would also enable association constants and timing of the integration reaction to be determined. However, because recombination between *attB* and *attP* sites was never detected without adding cell lysate to the recombination conditions, a more reliable assay was needed. This is the reason that made us move to develop an *in situ* recombination assay instead. The benefit of an *in situ* assay is that we do not need to be worry about the identification and purification of host encoded factors required for Int activity, as the conditions present in the cell are the conditions under which Int normally functions (bacterial cell cytoplasm). However, *in situ* assays provide us with less control over the amount and ratios of DNA substrates, proteins and enzymes that participate in the integration reaction. Furthermore, we need to extract the recombined DNA products at the end of the assay to analyse them.

The three compatible plasmids, *in situ* recombination assay enabled us to define, under the conditions tested, the minimal sequence length required to support integration events for both *attP* and *attB*. We determined that 427 bp of *attP* sequence was the smallest tested fragment that could support integration for

Φ24_B. Though this site is larger than *attP* for lambda phage (250 bp), it is similar to the minimal *attP* site identified for ICE_{clc} phage of *Pseudomonas putida* which is 450 bp (Miyazaki and van der Meer, 2013) or *attP* site identified for HP1 phage of *Haemophilus influenzae* which is 420 bp (Hauser and Scocca, 1992). Both of these *attP* sites, are also recognised by tyrosine recombinases and have similar size to the Φ24_B *attP* site (427 bp). However, we did not examine *attP* sequence lengths that were between 351-426 bp, therefore, further work is needed to more exactly define the Φ24_B *attP* site.

We found out that 93 bp of Φ24_B *attB* sequence was the smallest tested fragment that could support the integration of phage genome inside *E. coli*. So far, this is the biggest *attB* site for tyrosine integrase. This site is very much larger than *attB* for lambda phage (21 bp). (Lewis and Hatfull, 2003), reported that *attB* site needed for the phage L5 integrase activity is 29 bp inside *Mycobacterium smegmatis* chromosome, which is, also, smaller than the Φ24_B *attB* sequence. All other well-studied tyrosine integrases need an *attB* site less than that size (Fogg et al., 2014). On the other hand, integrases that belong to the serine family of recombinases use large *attB* sites to drive DNA recombination. For example, the minimal sequence length of *attB* site that R4 phage needs to introduce its genome inside *Streptomyces parvulus*, is 64 bp (Olivares et al., 2001). Nevertheless, we did not examine *attB* sequence lengths that were between 83-92 bp sizes. Therefore, further work is needed in future to find out the defined size of Φ24_B *attB* site needed for Φ24_B Int activity.

We demonstrated that host encoded IHF (integration host factor) is not necessary for $\Phi 24_B$ Int activity. Most tyrosine integrases can not mediate DNA recombination without the help of the IHF protein (Fogg et al., 2014). However, efficient integration derived by P22 phage integrase, which infects *Salmonella typhimurium*, can be detected even in the absence of IHF protein (Smith-Mungo et al., 1994). However, we also demonstrated that $\Phi 24_B$ Int requires some host-encoded accessory factor to drive DNA integration. DNA recombination derived by $\Phi 24_B$ Int in the *in vitro* recombination assay (chapter 3) did not take place without the addition of fresh, crude, cell extracts to the reaction test tube. Therefore, further work is needed here to find out what, within that cells lysate, is necessary for $\Phi 24_B$ Int activity.

The use of restriction-free cloning for most DNA cloning assays in this study provided a simple method to precisely insert a DNA fragment into its appropriate plasmid. In contrast to other methods that depend on DNA amplification, this method does not need the subsequent cloning of the amplified insert or restriction and phosphatase treatment of the plasmid vector. Instead, DNA amplification was used to fuse both insert and plasmid vector DNA. The cloning reaction is easily monitored and optimised (Shevchuk et al., 2004). In general, PCR outcomes were poor when the reaction primers fail to anneal or if they drove non-specific priming. These poor outcomes were identified when empty lanes were viewed in the analytical agarose gel. Sanger sequencing was done to confirm the identity of all products. The efficiency of the PCR reaction was optimized by

altering amplification parameters (annealing temperature, extension timings, the number of the PCR amplification cycles) and reagent concentrations (primers, templates, buffer ingredients).

The use of five or six compatible plasmids in the *in situ* recombination assay systems enabled us to characterise the preference and timing of integration in the secondary *attB* sites. We already knew that the lambdoid phage $\Phi 24_B$ had the ability to form multiple lysogens (able to integrate in several places within the *E. coli* genome). We also knew that more than one prophage could integrate into the *E. coli* chromosome and that these sequential integration events could actively promote the integration of another prophage (Fogg et al., 2007). There is an apparently intact lambda-like immunity region encoded within the $\Phi 24_B$ phage genome (Koudelka et al., 2004, Fattah et al., 2000), with equivalent repressor binding affinities, however, this region does not provide the host cell immunity to integration of a newly infecting phage in the manner in which lambda prophage provides immunity to superinfection. The $\Phi 24_B$ immunity region differs in organization from bacteriophage lambda. The $\Phi 24_B$ repressor CI binds to two operators only in pL rather than three, as in lambda. However, this does not affect the lysogen stability or repression of pL promoter transcription in the phage $\Phi 24_B$ and related Shiga toxin bacteriophages (Koudelka et al., 2004). Furthermore, the $\Phi 24_B$ Int is located the same genomic region as in the lambda genome, but the transcription of $\Phi 24_B$ Int lies in an inverted orientation, and that makes integrase expression outside the control of the global repressor protein CI. Instead $\Phi 24_B$ Int

expression is constitutive during stable lysogens, leading to accumulation of integrase within the host cytoplasm. The build up of active Int in the cytoplasm is the reason behind the raised frequency of prophage integration with superinfection while the orientation of Xis under the control of P_L enables excision of the prophage(s) still linked with induction (Fogg et al., 2011).

The integration assay run over various time points with cells harbouring either five or six compatible plasmids, showed *in situ* recombination frequencies five minutes after the induction of integrase expression into the most preferred secondary bacterial site *attB_{yfnr}* utilised $\sim 10^4$ fold fewer times than *attB₁*. The difference in utilisation differed only by a factor of the order of 10^3 fold less than *attB_{ints}* 14 hrs post expression of Int. The reason behind increased recombination frequency is likely because the increase of Int production from its gene during the time course experiment from its starting point (Fogg et al., 2011). Usage of *attB_{yfnr}* for recombination occurs at a much lower frequency than it does for *attB_{ints}*. This appears to be due to the fact that *attB_{yfnr}* shares only 62.5% identity to *attP* sequence. This idea is supported with the relationship between *attB* identity to *attP* and the usage of the *attB* site for recombination. The order of usage is *attB_{ints}*, *attB_{yfnr}*, *attB_{sgcA}*, and *attB_{yfbO}* at 100, 62.5, 52 and 50% identity to *attP*, respectively.

6.2 Future Work

6.2.1 Synthetic Recombination Vectors

Synthetic biology takes a ground-up approach in the genetic engineering of cellular systems capable of build novel biomolecule components, network and pathways, and to used these constructs rewire and reprogram biological systems (Khalil and Collins, 2010). The main goal of synthetic biology is to use tools to make it easier to engineer and preform increasingly complex biological systems when needed. The aim of this future work section is to describe the use of synthetic biology to exhaustively determine the required DNA nucleotides for the by $\Phi 24_B$ Int activity.

6.2.1.1 Integration and DNA homology

The importance of homology in DNA recombination driven by site-specific recombinases has been discussed in several previous studies. Some studies suggested that the homology of the DNA strands in recombination plays a role in the formation of the synapsis between to the strands to be recombined and HJ formation but not in the DNA strand migration or resolution of the HJ. They proposed that the requirement of DNA homology affects the DNA interaction during synapsis that preformed the creation of a four-strands helix (Kikuchi and Nash, 1979, Nash and Pollock, 1983). Nevertheless, in reactions promoted by lambda integrase, *attB* and *attP* sites that shared an overlap region with mismatches (3-4 bp) on the right side of its sequence were able to produce

uncompleted HJ intermediates as only one pair of DNA strand exchanges could generate a joint molecule of Holiday structure type (Fig.6.1) (Kitts and Nash, 1987). Experiments carried out with DNA substrates composed of half-*attP* sites comprised of either P or P' arm sequences and the heterologous crossover region, demonstrates that sequence mismatch did not affect the production of the DNA synapsis production or even the initial strand exchange event (Nunes-Düby et al., 1989).

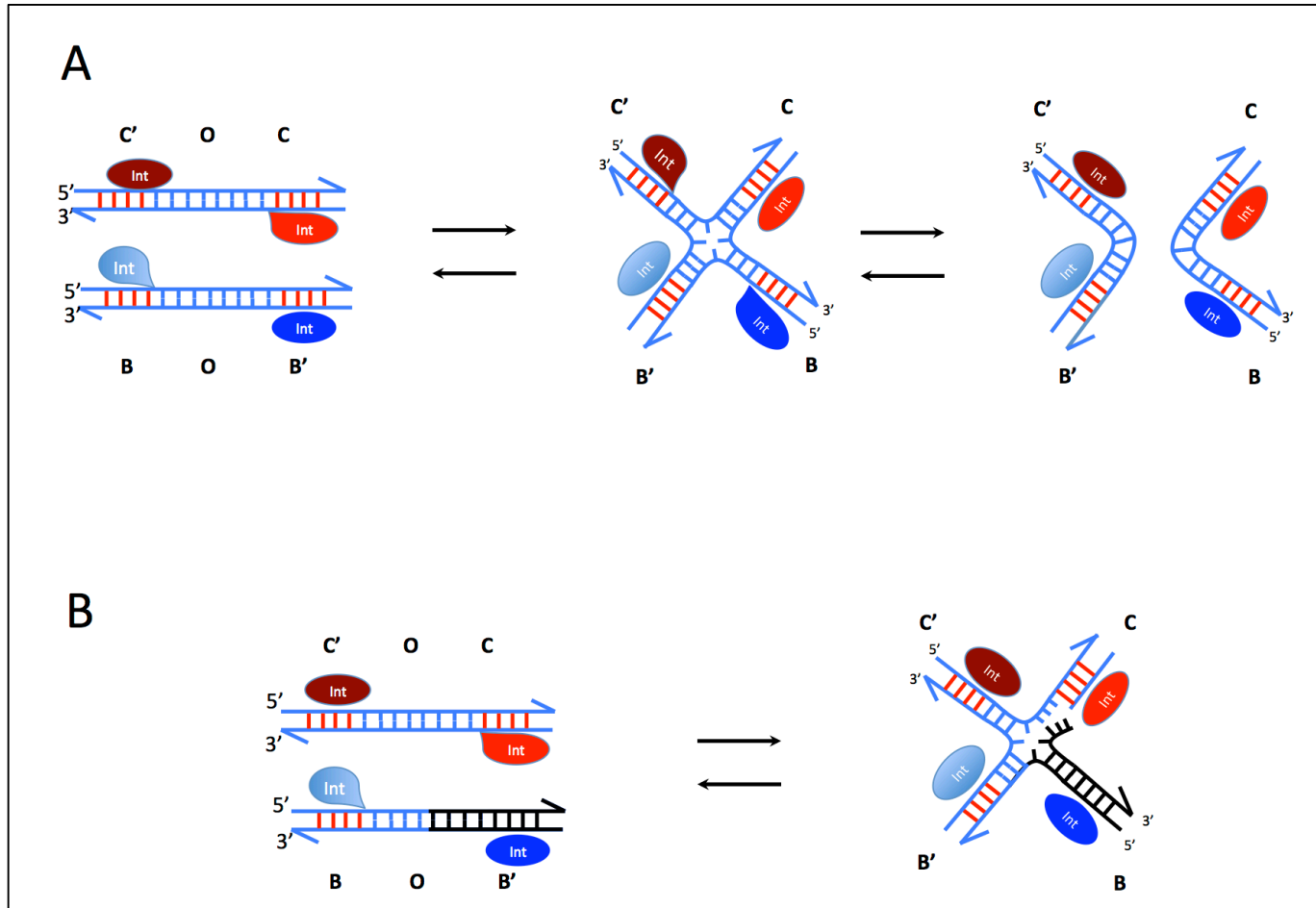


Figure 6.1. The effect of DNA sequence homology in the lambda *attB* and *attP* site recombination outcome mediated by lambda *Int*. A. when both *att* sites share an overlap region with complete matching sequence. The outcome in this case is the hybrid attL and attR sites. B. *attB* and *attP* sites that shared an overlap region with mismatches (Black DNA sequence) on the right side of attB sequence are able to produce uncompleted HJ intermediates as only one pair of DNA strand exchanges could generate a joint molecule of holiday structure type. The direction of resolution of the lambda *Int*-mediated HJs that has mismatches on the right side of the crossover region is backward to the parental products.

The outcome of each recombination event mediated by site-specific recombinases appears to be determined by the location of the sequence mismatch. In recombination mediated by lambda Int, HJ were observed to accumulate when mismatches were located in the right side of the core crossover region in *attB* (Fig.6.1), however the results were different when the mismatches occurred on the left hand side of the core crossover region where the first strand exchange creates HJ (Kitts and Nash, 1987). The direction of resolution of the lambda Int-mediated HJs that has mismatches on the right side of the crossover region is backward to the parental products (de Massy et al., 1989).

The role of sequence mismatch in the order of strand exchange governs the choice of secondary recombination sites into which phage lambda integrates. The first three nucleotides on the left side of the overlap region in all of the secondary *attB* sites are conserved. The remaining bases to the right are not conserved, probably because during phage integration into a secondary site, the HJ intermediates that are produced by first strand exchange are resolved by chromosome replication or by RuvC instead of Int (Rutkai et al., 2003).

If we examine the alignment of the $\Phi 24_B$ primary and secondary bacterial attachment sites (Fig 5.1 Chapter 5), we can clearly recognized that the bases on the right side of crossover core in the secondary *attB* sites contain more sequence mismatches than are found to the left of the cross over core. If we flip the crossover region orientation of *attP* to the *attB* sites and/or do some modification in the sequence bases, we might see

similar recombination outcomes to those described for lambda phage, above. These sequence manipulations would define our understanding of the impact of sequence mismatches between *attP* and *attB* sites in the recognition and integration of $\Phi 24_B$ phage genome in more than five different attB sites, as well as defining the nucleotides that are essential for $\Phi 24_B$ Int to mediate integration.

6.2.1.2 Gibson Assembly

Scientists have become able to join DNA sequences, for nearly four decades. This allows them to produce DNA segments that are not present in nature. The process of cutting and rejoining DNA sequences is called recombinant DNA technology. Recombinant technologies were initiated after the discovery of restriction enzymes and DNA ligase (Gellert, 1967, Smith and Wilcox, 1970). In 2009, Daniel Gibson and his group developed an *in vitro* recombination system that differed from the older DNA cutting and ligation methods and involved ligating DNA sequences a single isothermal reaction (Gibson et al., 2009) (Figure 6.2). The Gibson system requires a few enzymes to mediate the DNA combination reaction with minor manipulation. Three enzymes activities are required in Gibson reaction: an exonuclease (like 5' T5 exonuclease), a DNA polymerase (like Phusion DNA polymerase, Biolabs) and a DNA ligase (like *Taq* ligase, Biolabs). Up to 15 DNA fragments can be simultaneously combined. The combination between two DNA fragments require them to have ~20-40 bp overlap at there adjacent ends. These activities are sufficient to join DNA fragments as large as 500 kb. All required enzymes and reagents are commercially available.

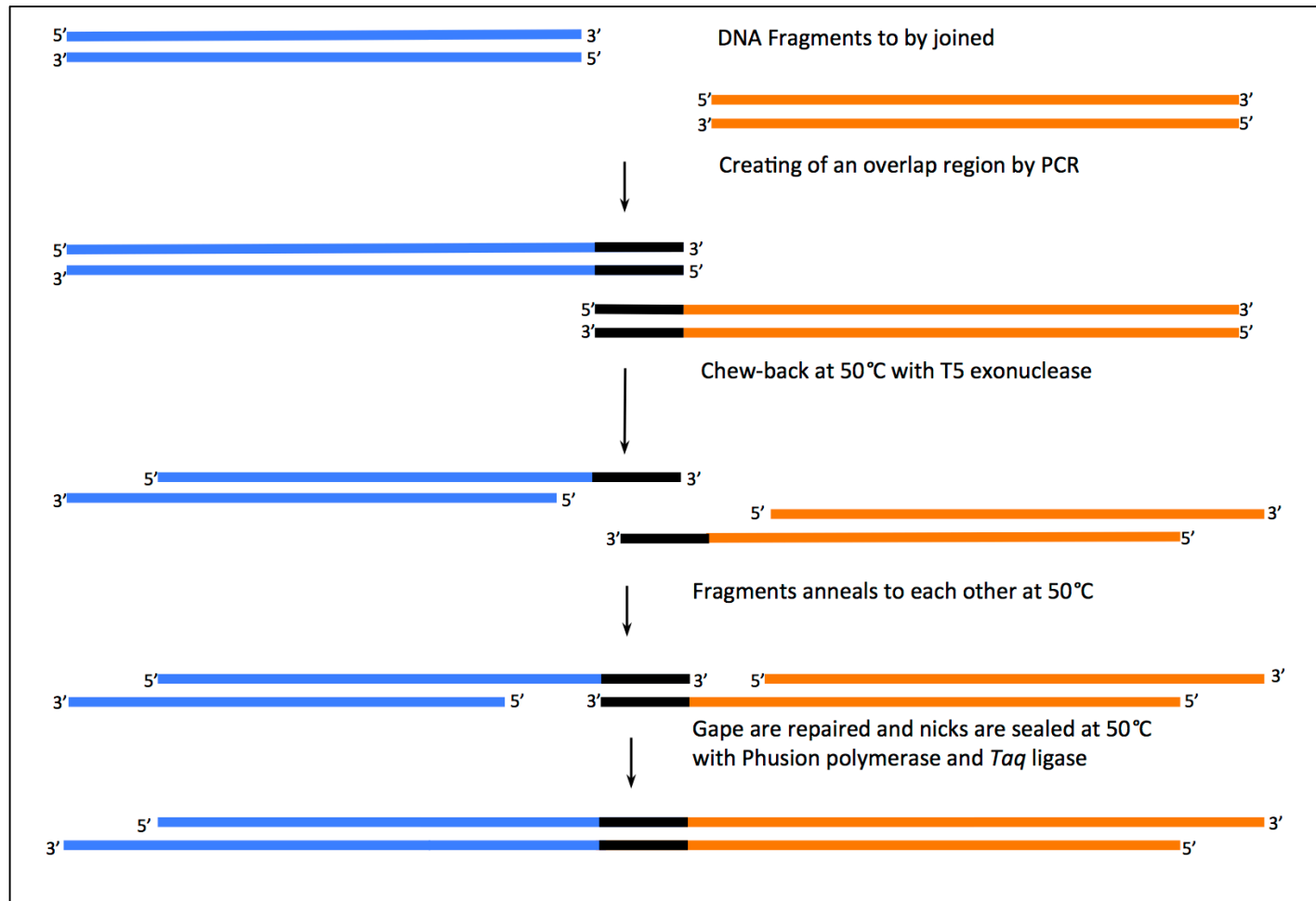


Figure 6.2. Gibson assembly overview. A short region of homology is created on both of the DNA fragments (Black region) so that there is a region of complementarity shared by both DNA fragments. The T5 exonuclease chews back the 5' ends in both fragments, which allows the two different fragments to anneal to each other. After that the Phusion DNA polymerase closes all the gaps, and *Taq* DNA ligase finishes the process by healing the remaining nicks in the sequences (Gibson et al., 2009).

An enzymatic master mix is produced from ('Isothermal Reaction (Gibson Assembly) Master Mix,' 2017). The cloning reaction is performed by the incubation of the master mix with DNA fragments sharing regions of complementarity at their ends at 50 °C for a few short minutes (time depend on the desire outcome and the enzymes and reagents requirements), which simplifies the creation of biological systems using small constituent parts.

Exonucleases, like the 5' T5, are enzymes that cleave DNA fragments one at time from the 5' ends of a polynucleotides chain. The exonuclease activity (5' T5 exonuclease, Epicentre) results in the production of single stranded regions at the end of the desired DNA fragments (Fig. 6.2), which can anneal to each other for the next step in the cloning reaction. The activity of DNA polymerase (Phusion DNA polymerase, Biolabs) then adds nucleotides to fill in the gaps. Finally the action of DNA ligase (*Taq* DNA ligase, Biolabs) joins the DNA sequences and resolves any nicks in the DNA products (Fig. 6.2)(Gibson et al., 2009) .

There are two kinds of Gibson assembly approaches: a one step assay or a two step assay. The one step assembly, is used for assembly of up to 5 different DNA fragments using a single-step, isothermal process. The DNA fragments are mixed with the enzyme master mix before the entire mixture is incubated at 50 °C for one hour. For the construction of more complicated DNA products, composed of up to 15 DNA fragments, a two step Gibson approach is used. This process

requires separate additions of the relevant enzymes rather than use of a master mix; the first enzyme addition is for exonuclease during which the 5' end is removed and the annealing step occurs. Then the DNA polymerase and DNA ligase are added. Each step has different optimal incubation temperatures to perform the DNA assembly (Gibson, 2011).

6.2.1.3 BioBricks

BioBricks are DNA sequences that work a bit like wooden or plastic building blocks. BioBricks are used to assemble larger, synthetic, biological circuits. These building parts have a defined structure and functions, which enable them to be incorporated into living cells to artificially construct new biological systems. There are many examples of these parts: ribosomal binding sites, promoters, terminators and coding sequences. These functional parts can be combined in novel combinations to build a *Device*, comprised of complementary BioBrick parts. A collection of devices can be linked to perform as a higher biological system. The reliability of a synthetic higher biological system is improved because of the ability to test each individual biological part and each device (Shetty et al., 2008). Every BioBrick part is currently a physical DNA sequence on a circular plasmid that is stored and distributed by the Registry of Biological Parts (<http://www.partsregistry.org>) (Sleight et al., 2010).

In order to perform these biological systems, several BioBricks assembly standards can be used. The first BioBricks assembly standard was reported in 2003 by Tom Knight (Shetty et al., 2008); it was called “standard sequence assembly of BioBricks”. However, since 2003, this very first assembly method has gone out of favour and many other assembly techniques have been reported (Røkke et al., 2014).

The ability of the standard assembly method to join two BioBricks depends on their sequences via the action of restriction endonuclease enzyme digestion and ligation (Canton et al., 2008). The key innovation of the BioBrick assembly standard is that two BioBrick parts can be assembled together, and the resulting product itself can still be considered to be a BioBrick part ready to be joined with any other BioBrick part. The BioBrick assembly standard enables a collection of compatible biological parts from different biological sources.

In contrast to the traditional *ad hoc* molecular approaches, this assembly is amenable to optimization and most importantly, automation (Shetty et al., 2008). All BioBricks components are flanked by short functional sequences known as prefixes and suffixes. The prefix and suffix sequences are useful for joining together two different BioBrick parts, as they contain restriction endonuclease digestion sites. Different assembly standards have been developed. Each proposed assembly has its specific prefix and suffix (with a unique digestion site) flanking the BioBricks. The main goal of most assembly standards is to make BioBrick

fusions of protein coding domains possible. Stability and degradation rates of the fusion proteins are issues that have led to more recently proposed assembly standards (Røkke et al., 2014).

6.2.1.4 BioBrick Standard Assembly

BioBrick Standard Assembly or Assembly standard 10, was proposed by Tom Knight in 2007 (Røkke et al., 2014). It is the most commonly used assembly standard. Each BioBrick part is actually a DNA sequence carried by a circular plasmid. The first step toward the BioBrick standard is the introduction of the prefix and suffix sequences with their specific restriction enzymes sites (Voigt, 2011). In this assembly standard, BioBricks have one standardized suffix while one of two types of prefix can be used. Which type of prefix depends on whether or not the BioBrick is a protein-coding sequence or not. Both prefixes have sites that can be recognized and cleaved by restriction endonucleases *EcoR* I and *Not* I. The non-protein coding prefix has, in addition to these two sites, a site utilizable by *Xba* I. The *Xba* I site present in the protein-coding prefix is fused with a protein coding sequence starting with ATG (Røkke et al., 2014). Restriction endonuclease sites that are present in the suffix sequence are recognized by *Spe* I, *Not* I and *Pst* I enzymes (Radeck et al., 2013). The common method to combine two BioBrick parts following the BioBrick Standard Assembly, is to digest the first BioBrick using the restriction enzymes *EcoR* I and *Spe* I. The second BioBrick is digested with *EcoR* I and *Xba* I (Fig.6.3)(Sleight et al., 2010).

The two digested BioBricks then both possess sticky ends that were created by *EcoR* I and can be easily joined together to restore a recognisable *EcoR* I site. On the other hand, the sticky ends that were created by *Xba* I and *Spe* I are compatible and can be ligated together to produce a ligated DNA molecule that will be a fusion of the two restriction sites (Fig.6.3) (Liu et al., 2014). This new DNA sequence cannot be recognized by *Spe* I or *Xba* I. This new sequence, now called a scar, will be composed of the bases TACTAGAG in the cases when the second BioBrick is not a protein coding part or it will be TACTAG if that BioBrick is a protein-coding region (Røkke et al., 2014).

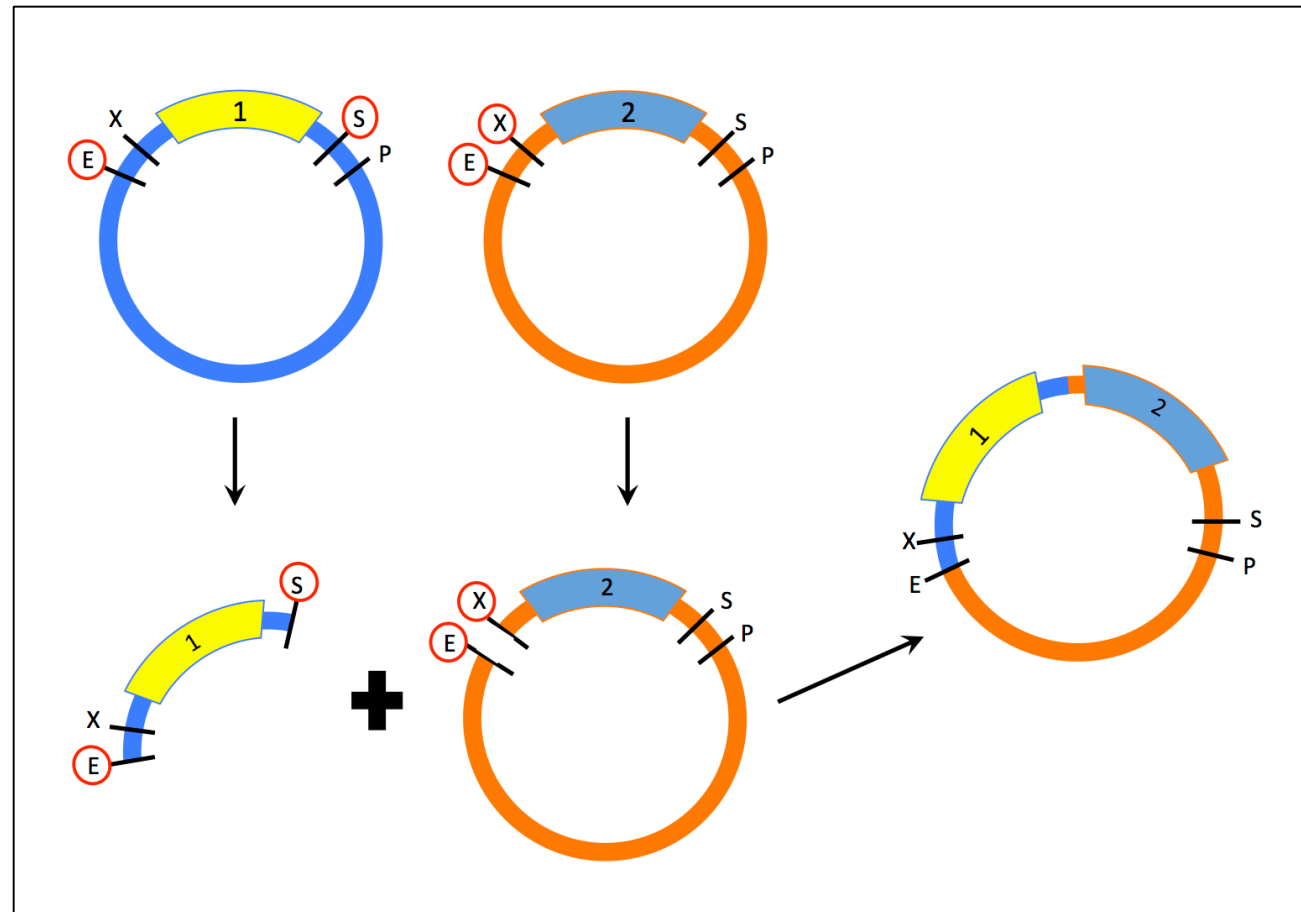


Figure 6.3. A schematic diagram depicting the assembling of two BioBricks using BioBrick Standard Assembly. The restriction enzyme that used to digest the first part (1) are *EcoR* I and *Spe* I, and that used for the second part (2) are *EcoR* I and *Xba* I. The digested DNA molecules are then ligated, producing a new BioBrick made up of the two initial ones. Here E, X, S and P denotes *EcoR* I, *Xba* I, *Spe* I, and *Pst* I, respectively.

This commonly used assembly was an example of BioBricks assembly standard approaches; other approaches are BglBricks assembly standard (Lee et al., 2011), Silver (Biofusion) standard and Freiburg standard (Røkke et al., 2014). The difference among them is in the the prefix and suffix sequences and the restriction sites within these sequences.

6.2.2. Proof of concept: synthetic recombination vector assay

In order to create a synthetic recombination assay and identify the essential nucleotides in the *attB* sites required by $\Phi 24_B$ Int to promote the recombination, we have design a vector that harbors both the *attB_{intS}* and *attP* sites facing in opposite directions. These sites lie within a synthetically produced fragment of DNA that is 5271 bp long. The full description of this synthetically derived DNA is *attP* (427 bp) derived from $\Phi 24_B$, GFP promoter sequence, the full gene encoding the β -galactosidase enzyme (*lacZ*), T7 terminator sequence, *attB_{intS}* (124 bp) from *E. coli* strain MC1061 and a promoterless green fluorescent reporter gene (GFP) (Fig. 6.4). The GFP reporter is placed within the vector so that it is expressed following a recombination event. The plasmid backbone for this synthetic vector is the high copy number plasmid, pCDFDuet, which has a spectinomycin resistant gene as a selectable marker. This work was done in cooperation with Gene Mill within the Centre for Genomic Research at the University of Liverpool.

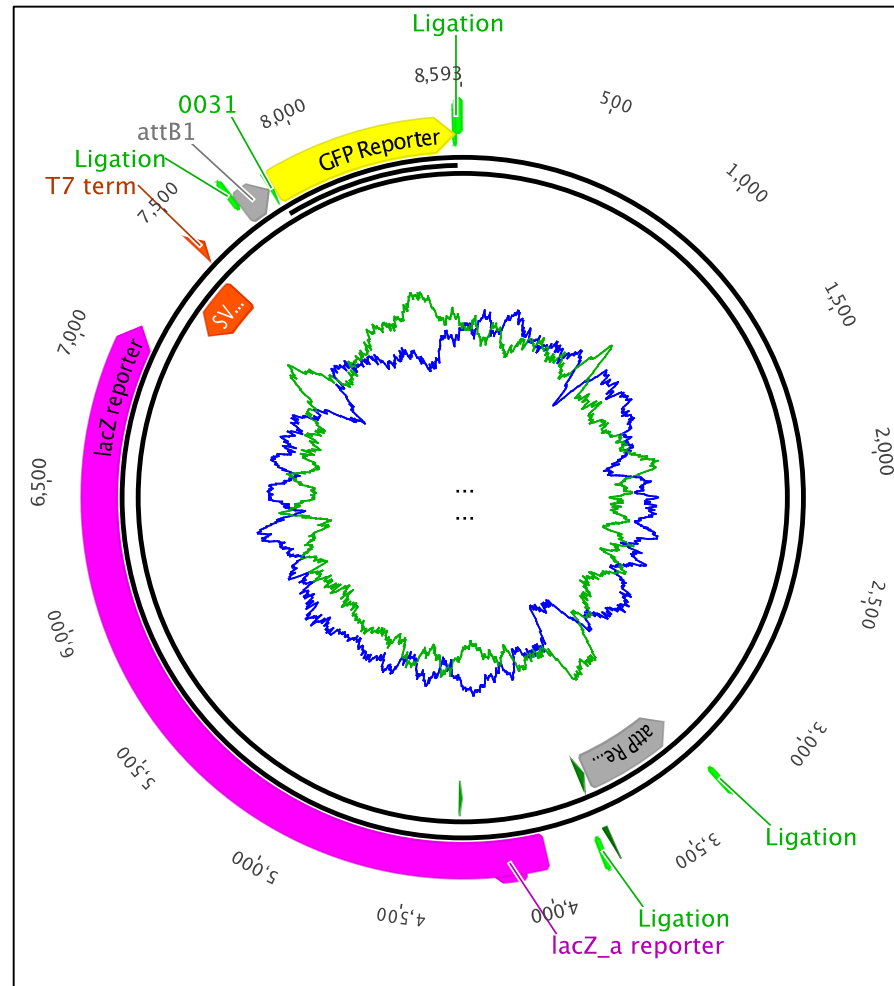


Figure 6.4. The synthesized integration vector. The size of the entire vector is 8593 bp. *attB_{ints}* in this version is in forward direction and its size is 124 bp (highlighted with gray color (attB1)). *attP* shown here is in the reverse direction and is 427 bp (highlighted with gray color). GFP promoter sequence presented with dark green color, the *lacZ* gene with its own promoter highlighted with pink color, T7 terminator sequence presented with red color, and GFP gene highlighted with yellow color. The ligation sites are labeled with light green. The *lacZ* gene and the pCDFDuet plasmid provide ample flexibility for recombination to occur intramolecularly.

There are two reasons behind the use of the *lacZ* gene in the synthetic construct: One, to provide enough flexibility in the DNA plasmid to enable intramolecular recombination events; and two, to be able to screen for blue colonies after the transformation of the vector into cells when they grow on a media containing X-gal (Fig. 6.2).

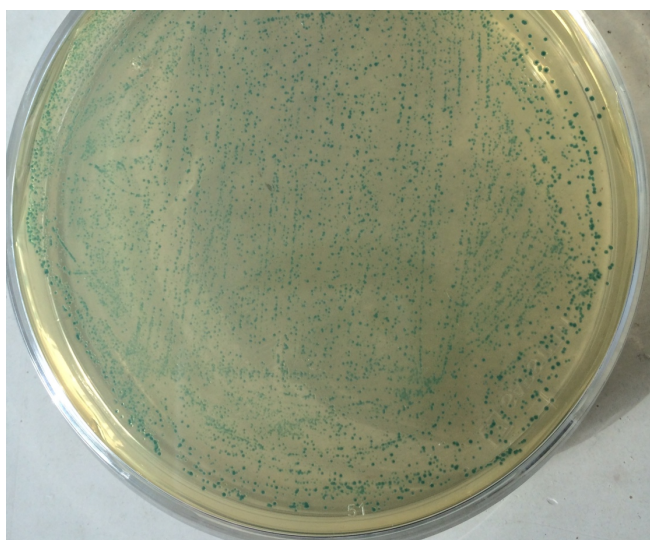


Figure 6.5. Bacterial growth of Top10 *E. coli* cells following transformation of the integration vector.

Since the start point of this work, seven vectors, including a positive control, were created in order to measure recombination events through the production of green fluorescence cells (Table 6.1). After the production of each plasmid, it was transformed into Top10 *E. coli* competent cells that already harbored the Pint plasmid. Cells were grown overnight for *in situ* recombination assays (see 4.2.2. *In situ* recombination assay, Chapter 4). After each completed assay, samples were taken for examination by fluorescence microscopy.

Table 6.1. The recombination vectors that we created for synthetic recombination derived by $\Phi 24_B$ Int.

| Vector Version | Vector details | The recombination assay result |
|------------------|---|--|
| V1 | Int test V1, attP-Rev., attB –Fwd, Wk GFP RPS | Recombination resulted in the production of two DNA molecules and no fluorescence. |
| V2 | Int test V2, attP-Fwd., attB-Fwd, Wk GFP RPS | Recombination resulted in the production of two DNA molecules and no fluorescence. |
| V3 | Int test V3, attP-Fwd., attB-Fwd, Str GFP RPS | Recombination resulted in the production of two DNA molecules and no fluorescence. |
| V4 | Int test V4, attP-Fwd., attB-Fwd, Str GFP RPS | Recombination resulted in the production of two DNA molecules and no fluorescence. |
| V5 | Int test V5, attP-Rev., attB-Fwd, Med GFP RPS | Recombination resulted in the production of two DNA molecules and no fluorescence. |
| V6 | Int test V6, attP-Rev., attB-Fwd, Str GFP RPS | Recombination resulted in the production of two DNA molecules and no fluorescence. |
| Positive control | Int positive control V., attR., attL, Str GFP RPS | This gave very nice fluorescent cells. |

cell after recombination. DNA Recombination was confirmed following recovery of the plasmid DNA from the cells and a PCR reaction using *attRL* primers. The identity of the amplified bands was confirmed by Sanger sequencing. Furthermore, the recovered plasmid DNA was transformed into new Top10 cells, grown on agar media containing spectinomycin (give final concentration) and subjected to restriction digest with *Apa* I to confirm if it was a single molecule of the right size (state correct size here) (Fig. 6.3).

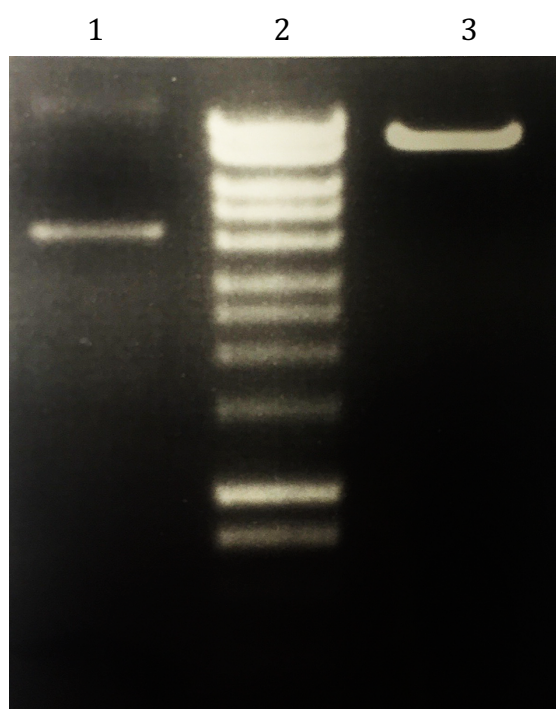


Figure 6.6. Recombination product following plasmid harvest, restriction enzyme digestion and electrophoresis gel run. Lane 1) Result of the vector recombination that resulted in two separate, DNA molecules, each molecule's size ~ 4300 bp during recombination. Lane 2) MW, Hyperladder 1 kb (Bioline) . Lane 3) Result of the vector recombination that resulted in one DNA molecule of the same of as the original vector (8593 bp).

All of our synthetic vectors were capable of supporting recombination, however we were unable to detect fluorescence from the recombination product. The result was either a recombination producing a single recombined plasmid that should enable GFP expression, or recombination producing two circular DNA molecules (Fig. 6.8) that produce white and blue colonies when the recombination products were extracted from the cells in the *in situ* recombination assay and transformed into new competent cells (Fig. 6.7 A&B).

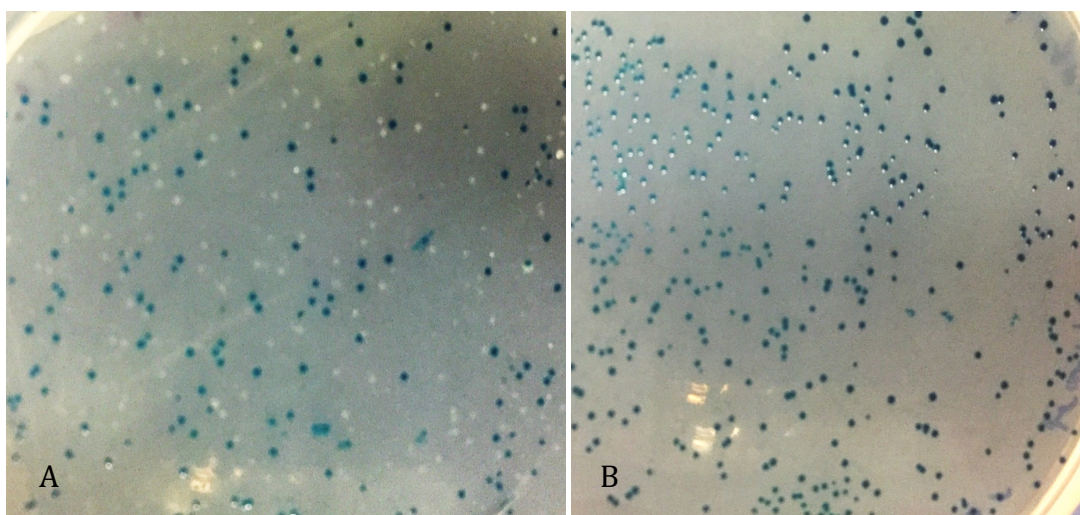


Figure 6.7. Bacterial colonies that resulted after the transformation of the purified recombined vector to New Top10 competent cells. A: Result of the vector recombination that gave two separate DNA molecules. One of these molecules has the antibiotic marker gene and the other has the LacZ reported gene, the white colonies are for cells harboring the marker gene DNA molecules only while the blue cells have both of them as the vector backbone is the high copy number pCDFDuet plasmid. B: These unique blue colonies are performed after the transformation of the positive recombined vector that gave one DNA molecule.

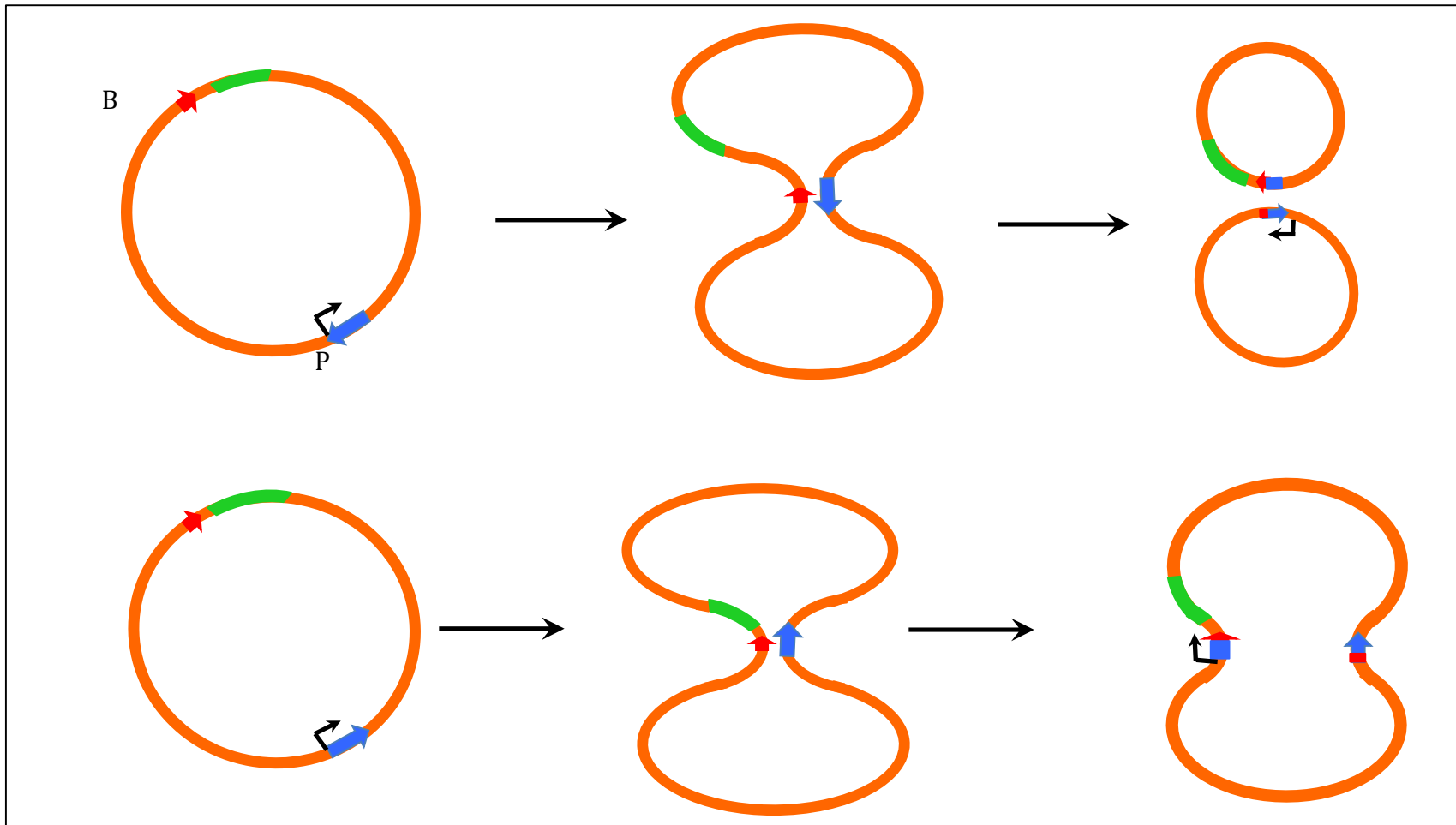


Figure 6.8. A schematic diagram explains the recombination events between attP (blue arrows) and attB (Red arrows) sites within the synthesized integration vector. The results that we got was either: A , Recombination gave two separate circular DNA molecules, B, or Recombination resulted to a one molecule recombined vector, and this is the right way of recombination that should allow the GFP transcription.

Nevertheless, we were so closed to have fluorescent cells from the recombination as the positive version of the synthetic vector gave so nice fluorescent cells (Pic, 6.4). Farther work is needed to find out why the GFP gene does not transcript properly even with positive recombination. Using this assay for determine the required enucleated for the by $\Phi 24_B$ Int is so less time consuming and so less cost.

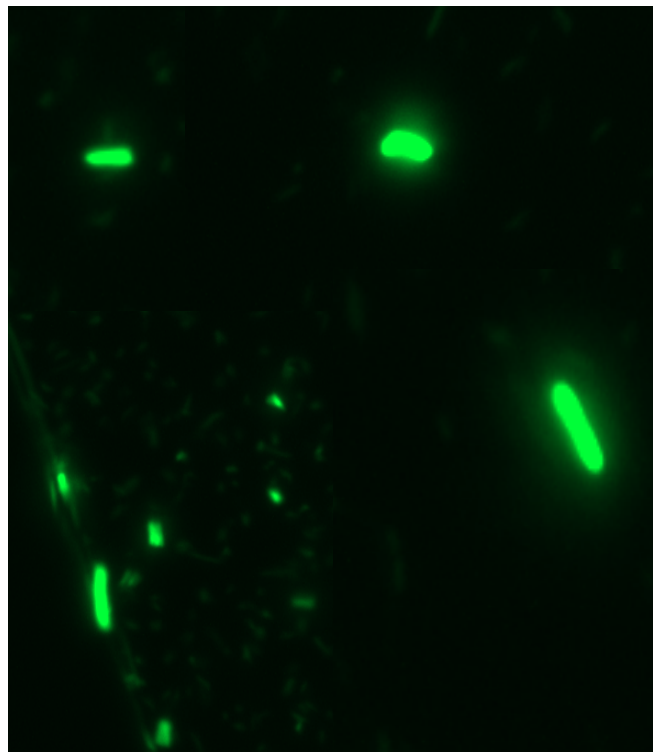


Figure 6.9. Positive control vector demonstrating the ability to capture integration events. The successful production of this vector will can be followed by the synthesis of directed mutations in all 4 *attB* sites to identify the nucleotide sequences that are essential for recombination.

6.2.3 Necessary Integration Host Factors

$\Phi 24_B$ Int is tyrosine recombinase, most of tyrosine integrases can not drive the phage genome recombination inside the host chromosome without assistance of the host encoded accessory proteins.

The aim of this future work is to identify the necessary host factor or factors for $\Phi 24_B$ Int directed integration.

6.2.3.1 Protein fractionation

Protein fractionation is a process or series of processes intended to isolate a single or multiple types of proteins from a complex mixture of proteins. The most common approach used for protein fractionation is column chromatography fractionation. In which the solution of the proteins mixture is passed through a column containing a porous solid matrix. This system uses different extents to retard different proteins by their interaction with the matrix, and they can be collected separately as they flow out of the bottom of the column. Proteins can be separated depends on their size (gel-filtration chromatography), charge (ion-exchange chromatography), hydrophobicity (hydrophobic chromatography), or their ability to bind to particular small molecules or to other macromolecules (affinity chromatography) (Wilson and Hunt, 2002)

The size and subunit composition of a protein can be determined by SDS polyacrylamide-gel electrophoresis (Zhang et al., 2008a).

Other techniques can be used in protein fractionation are homogenization (Schmidt et al., 2012), centrifugation (Siegel and Monty, 1966), fractionation by precipitation using ammonium sulfate (Wingfield, 2001) and ultra filtration (Bonnaillie et al., 2014).

6.2.3.2 Proof of concepts

Lambda integrase is responsible for the direction of lambda phage genome inside a unique site in the bacterial chromosome. However, this protein cannot promote its function without presence of the bacterial encoded DNA binding protein IHF or Integration Host Factor (Craig and Nash, 1984). As $\Phi 24_B$ Bacteriophage is lambdoid family member, we considered IHF would be so important for $\Phi 24_B$ Int for directed integration, as most of other family members. Therefore, we performed an *in vitro* recombination assay with purified IHF protein and Int only (see chapter 3). This assay had not given any positive recombination until we added a cells crude extract to the reaction tube. Furthermore, to confirm the importance of IHF for $\Phi 24_B$ Int function *in situ* recombination assay was performed in the JW1702-1 *E coli* cells that have a deletion mutation of the *himA* gene inside its genome (one of two IHF genes) (see chapter 4). The results showed that recombination was actually performed. The results of all these assays suggested there is another bacterial host accessory factor(s) needed for the $\Phi 24_B$ integrase to function as a site-specific recombinase. Therefore, further work is needed to be performed to find out what is the necessary host encoded protein needed for $\Phi 24_B$ Int to promote the DNA recombination.

CHAPTER 7.: REFERENCES

- Abremski, K. and Gottesman, S. (1982) 'Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination', *J Biol Chem*, 257(16), pp. 9658-62.
- Abremski, K. and Hoess, R. (1984) 'Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein', *J Biol Chem*, 259(3), pp. 1509-14.
- Allison, H. E. (2007) 'Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens', *Future Microbiol*, 2(2), pp. 165-74.
- Allison, H. E., Sergeant, M. J., James, C. E., Saunders, J. R., Smith, D. L., Sharp, R. J., Marks, T. S. and McCarthy, A. J. (2003) 'Immunity profiles of wild-type and recombinant shiga-like toxin-encoding bacteriophages and characterization of novel double lysogens', *Infect Immun*, 71(6), pp. 3409-18.
- Andrews, B. J., Proteau, G. A., Beatty, L. G. and Sadowski, P. D. (1985) 'The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences', *Cell*, 40(4), pp. 795-803.
- Atlung, T., Nielsen, A., Rasmussen, L. J., Nellemann, L. J. and Holm, F. (1991) 'A versatile method for integration of genes and gene fusions into the lambda attachment site of Escherichia coli', *Gene*, 107(1), pp. 11-7.
- Austin, S., Ziese, M. and Sternberg, N. (1981) 'A novel role for site-specific recombination in maintenance of bacterial replicons', *Cell*, 25(3), pp. 729-36.
- Arfin, S. M., Long, A. D., Ito, E. T., Toller, L., Riehle, M. M., Paegle, E. S. and Hatfield, G. W. (2000) 'Global gene expression profiling in Escherichia coli K12. The effects of integration host factor', *J Biol Chem*, 275(38), pp. 29672-84.
- Azam, T. A. and Ishihama, A. (1999) 'Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity', *J Biol Chem*, 274(46), pp. 33105-13.
- Azaro, M. A. and Landy, A. (1997) 'The isomeric preference of Holliday junctions influences resolution bias by lambda integrase', *EMBO J*, 16(12), pp. 3744-55.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. and Mori, H. (2006) 'Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection', *Mol Syst Biol*, 2, pp. 2006.0008.
- Belteki, G., Gertsenstein, M., Ow, D. W. and Nagy, A. (2003) 'Site-specific cassette exchange and germline transmission with mouse ES cells expressing ϕ C31 integrase', *Nature Biotechnology* 21, pp. 321 - 324.

- Berry, J., Rajaure, M., Pang, T. and Young, R. (2012) 'The spanin complex is essential for lambda lysis', *J Bacteriol*, 194(20), pp. 5667-74.
- Berry, J., Summer, E. J., Struck, D. K. and Young, R. (2008) 'The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes', *Mol Microbiol*, 70(2), pp. 341-51.
- Bharati, K. and Ganguly, N. K. (2011) 'Cholera toxin: a paradigm of a multifunctional protein', *Indian J Med Res*, 133, pp. 179-87.
- Bibb, L. A., Hancox, M. I. and Hatfull, G. F. (2005) 'Integration and excision by the large serine recombinase phiRv1 integrase', *Mol Microbiol*, 55(6), pp. 1896-910.
- Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. and Ellenberger, T. (2005) 'A structural basis for allosteric control of DNA recombination by lambda integrase', *Nature*, 435(7045), pp. 1059-66.
- Bliska, J. B. and Cozzarelli, N. R. (1987) 'Use of site-specific recombination as a probe of DNA structure and metabolism in vivo', *J Mol Biol*, 194(2), pp. 205-18.
- Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. and Falkow, S. (1977) 'Construction and characterization of new cloning vehicles. II. A multipurpose cloning system', *Gene*, 2(2), pp. 95-113.
- Bonnaillie, L. M., Qi, P., Wickham, E. and Tomasula, P. M. (2014) 'Enrichment and Purification of Casein Glycomacropeptide from Whey Protein Isolate Using Supercritical Carbon Dioxide Processing and Membrane Ultrafiltration', *Foods*, 3(1), pp. 94-109.
- Botstein, D. (1980) 'A theory of modular evolution for bacteriophages', *Ann N Y Acad Sci*, 354, pp. 484-90.
- Branda, C. S. and Dymecki, S. M. (2004) 'Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice', *Dev Cell*, 6(1), pp. 7-28.
- Bradley, D. E. (1967) 'Ultrastructure of bacteriophage and bacteriocins', *Bacteriol Rev*, 31(4), pp. 230-314.
- Breitbart, M. and Rohwer, F. (2005) 'Here a virus, there a virus, everywhere the same virus?', *Trends Microbiol*, 13(6), pp. 278-84.
- Breüner, A., Brøndsted, L. and Hammer, K. (1999) 'Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1', *J Bacteriol*, 181(23), pp. 7291-7.
- Breüner, A., Brøndsted, L. and Hammer, K. (2001) 'Resolvase-like recombination performed by the TP901-1 integrase', *Microbiology*, 147(Pt 8), pp. 2051-63.
- Broach, J. R., Guarascio, V. R. and Jayaram, M. (1982) 'Recombination within the yeast plasmid 2mu circle is site-specific', *Cell*, 29(1), pp. 227-34.

- Bryksin, A. V. and Matsumura, I. (2010) 'Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids', *Biotechniques*, 48(6), pp. 463-5.
- Brüssow, H., Canchaya, C. and Hardt, W. D. (2004) 'Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion', *Microbiol Mol Biol Rev*, 68(3), pp. 560-602, table of contents.
- Burnet, F. M. (1934) 'The Bacteriophage, Biological reviews', *Cambridge Philosophical Society*, 9(3), pp. 332-350.
- Bustin, S. A. (2000) 'Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays', *J Mol Endocrinol*, 25(2), pp. 169-93.
- Calendar, R. (ed.) (2006) *The bacteriophages*. 2nd edn. Madison Avenue, New York, USA: Oxford University Press, Inc.
- Campbell, A. (2003) 'The future of bacteriophage biology', *Nat Rev Genet*, 4(6), pp. 471-7.
- Campbell, A. (1962) 'Episomes', *Advances in Genetics*, 11, pp. 101-145.
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. and Brüssow, H. (2003) 'Phage as agents of lateral gene transfer', *Curr Opin Microbiol*, 6(4), pp. 417-24.
- Canton, B., Labno, A. and Endy, D. (2008) 'Refinement and standardization of synthetic biological parts and devices', *Nat Biotechnol*, 26(7), pp. 787-93.
- Carter, J. B. and Saunders, V. A. (2007) *Virology : principles and applications*. Chichester: John Wiley.
- Cassell, G., Moision, R., Rabani, E. and Segall, A. (1999) 'The geometry of a synaptic intermediate in a pathway of bacteriophage lambda site-specific recombination', *Nucleic Acids Res*, 27(4), pp. 1145-51.
- Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. and Rice, P. A. (2000) 'Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping', *Mol Cell*, 6(4), pp. 885-97.
- Cho, E. H., Gumport, R. I. and Gardner, J. F. (2002) 'Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex', *J Bacteriol*, 184(18), pp. 5200-3.
- Christ, N. and Dröge, P. (2002) 'Genetic manipulation of mouse embryonic stem cells by mutant lambda integrase', *Genesis*, 32(3), pp. 203-8.
- Christiansen, B., Brøndsted, L., Vogensen, F. K. and Hammer, K. (1996) 'A resolvase-like protein is required for the site-specific integration of the temperate lactococcal bacteriophage TP901-1', *J Bacteriol*, 178(17), pp. 5164-73.
- Collier, R. J. (1975) 'Diphtheria toxin: mode of action and structure', *Bacteriol Rev*, 39(1), pp. 54-85.

- Combes, P., Till, R., Bee, S. and Smith, M. C. (2002) 'The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system', *J Bacteriol*, 184(20), pp. 5746-52.
- Court, D. L., Oppenheim, A. B. and Adhya, S. L. (2007) 'A new look at bacteriophage lambda genetic networks', *J Bacteriol*, 189(2), pp. 298-304.
- Craig, N. L. (1988) 'The mechanism of conservative site-specific recombination', *Annu Rev Genet*, 22, pp. 77-105.
- Craig, N. L. and Nash, H. A. (1984) 'E. coli integration host factor binds to specific sites in DNA', *Cell*, 39(3 Pt 2), pp. 707-16.
- Creuzburg, K., Recktenwald, J., Kuhle, V., Herold, S., Hensel, M. and Schmidt, H. (2005) 'The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein', *J Bacteriol*, 187(24), pp. 8494-8.
- de la Cruz, F. and Davies, J. (2000) 'Horizontal gene transfer and the origin of species: lessons from bacteria', *Trends Microbiol*, 8(3), pp. 128-33.
- de Massy, B., Dorgai, L. and Weisberg, R. A. (1989) 'Mutations of the phage lambda attachment site alter the directionality of resolution of Holliday structures', *EMBO J*, 8(5), pp. 1591-9.
- Dodd, I. B., Perkins, A. J., Tsemitsidis, D. and Egan, J. B. (2001) 'Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny', *Genes Dev*, 15(22), pp. 3013-22.
- Dodd, I. B., Shearwin, K. E., Perkins, A. J., Burr, T., Hochschild, A. and Egan, J. B. (2004) 'Cooperativity in long-range gene regulation by the lambda CI repressor', *Genes Dev*, 18(3), pp. 344-54.
- Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988) 'High efficiency transformation of E. coli by high voltage electroporation', *Nucleic Acids Res*, 16(13), pp. 6127-45.
- Dul, J. L. and Drexler, H. (1988) 'Transcription stimulates recombination. I. Specialized transduction of Escherichia coli by lambda trp phages', *Virology*, 162(2), pp. 466-70.
- Dymecki, S. M. (1996) 'Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice', *Proc Natl Acad Sci U S A*, 93(12), pp. 6191-6.
- Echols, H. and Green, L. (1979) 'Some properties of site-specific and general recombination inferred from int-initiated exchanges by bacteriophage lambda', *Genetics*, 93(2), pp. 297-307.
- Ennis, D. G., Fisher, B., Edmiston, S. and Mount, D. W. (1985) 'Dual role for Escherichia coli RecA protein in SOS mutagenesis', *Proc Natl Acad Sci U S A*, 82(10), pp. 3325-9.

- Enquist, L. W., Kikuchi, A. and Weisberg, R. A. (1979) 'The role of lambda integrase in integration and excision', *Cold Spring Harb Symp Quant Biol*, 43 Pt 2, pp. 1115-20.
- Esposito, D., Garvey, L. A. and Chakiath, C. S. (2009) 'Gateway cloning for protein expression', *Methods Mol Biol*, 498, pp. 31-54.
- Faruque, S. M. (2014) 'Role of phages in the epidemiology of cholera', *Curr Top Microbiol Immunol*, 379, pp. 165-80.
- Faruque, S. M., Rahman, M. M., Asadulghani, Nasirul Islam, K. M. and Mekalanos, J. J. (1999) 'Lysogenic conversion of environmental *Vibrio mimicus* strains by CTXPhi', *Infect Immun*, 67(11), pp. 5723-9.
- Fattah, K. R., Mizutani, S., Fattah, F. J., Matsushiro, A. and Sugino, Y. (2000) 'A comparative study of the immunity region of lambdoid phages including Shiga-toxin-converting phages: molecular basis for cross immunity', *Genes Genet Syst*, 75(5), pp. 223-32.
- Fauquet, C. M. and Fargette, D. (2005) 'International Committee on Taxonomy of Viruses and the 3,142 unassigned species', *Virology*, 2, pp. 64.
- Fien, K., Turck, A., Kang, I., Kielty, S., Wulff, D. L., McKenney, K. and Rosenberg, M. (1984) 'CII-dependent activation of the pRE promoter of coliphage lambda fused to the *Escherichia coli* galK gene', *Gene*, 32(1-2), pp. 141-50.
- Fogg, P. C., Colloms, S., Rosser, S., Stark, M. and Smith, M. C. (2014) 'New applications for phage integrases', *J Mol Biol*, 426(15), pp. 2703-16.
- Fogg, P. C., Gossage, S. M., Smith, D. L., Saunders, J. R., McCarthy, A. J. and Allison, H. E. (2007) 'Identification of multiple integration sites for Stx-phage Phi24B in the *Escherichia coli* genome, description of a novel integrase and evidence for a functional anti-repressor', *Microbiology*, 153(Pt 12), pp. 4098-110.
- Fogg, P. C., Rigden, D. J., Saunders, J. R., McCarthy, A. J. and Allison, H. E. (2011) 'Characterization of the relationship between integrase, excisionase and antirepressor activities associated with a superinfecting Shiga toxin encoding bacteriophage', *Nucleic Acids Res*, 39(6), pp. 2116-29.
- Fonville, N. C., Bates, D., Hastings, P. J., Hanawalt, P. C. and Rosenberg, S. M. (2010) 'Role of RecA and the SOS response in thymineless death in *Escherichia coli*', *PLoS Genet*, 6(3), pp. e1000865.
- Freifelder, D. and Kirschner, I. (1971) 'The formation of homoimmune double lysogens of phage lambda and the segregation of single lysogens from them', *Virology*, 44(3), pp. 633-7.
- Freuler, F., Stettler, T., Meyerhofer, M., Leder, L. and Mayr, L. M. (2008) 'Development of a novel Gateway-based vector system for efficient, multiparallel protein expression in *Escherichia coli*', *Protein Expr Purif*, 59(2), pp. 232-41.

- Friedman, D. I. and Court, D. L. (1995) 'Transcription antitermination: the lambda paradigm updated', *Mol Microbiol*, 18(2), pp. 191-200.
- Friedman, D. I. (1988) 'Integration host factor: a protein for all reasons', *Cell*, 55(4), pp. 545-54.
- Frumerie, C., Sylwan, L., Helleday, T., Yu, A. and Haggård-Ljungquist, E. (2008) 'Bacteriophage P2 integrase: another possible tool for site-specific recombination in eukaryotic cells', *J Appl Microbiol*, 105(1), pp. 290-9.
- Frumerie, C., Sylwan, L., Ahlgren-Berg, A. and Haggård-Ljungquist, E. (2005) 'Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox', *Virology*, 332(1), pp. 284-94.
- Gellert, M. (1967) 'Formation of covalent circles of lambda DNA by E. coli extracts', *Proc Natl Acad Sci U S A*, 57(1), pp. 148-55.
- Geue, L., Klare, S., Schnick, C., Mintel, B., Meyer, K. and Conraths, F. J. (2009) 'Analysis of the clonal relationship of serotype O26:H11 enterohemorrhagic Escherichia coli isolates from cattle', *Appl Environ Microbiol*, 75(21), pp. 6947-53.
- Ghosh, P., Kim, A. I. and Hatfull, G. F. (2003) 'The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB', *Mol Cell*, 12(5), pp. 1101-11.
- Ghosh, P., Wasil, L. R. and Hatfull, G. F. (2006) 'Control of phage Bxb1 excision by a novel recombination directionality factor', *PLoS Biol*, 4(6), pp. e186.
- Gibson, D. G. (2011) 'Enzymatic assembly of overlapping DNA fragments', *Methods Enzymol*, 498, pp. 349-61.
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. and Smith, H. O. (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nat Methods*, 6(5), pp. 343-5.
- Goodman, S. D., Velten, N. J., Gao, Q., Robinson, S. and Segall, A. M. (1999) 'In vitro selection of integration host factor binding sites', *J Bacteriol*, 181(10), pp. 3246-55.
- Griffiths, A. J. F. (2000) *An introduction to genetic analysis*. 7th ed. edn. New York: W.H. Freeman.
- Groman, N. B. (1984) 'Conversion by corynephages and its role in the natural history of diphtheria', *J Hyg (Lond)*, 93(3), pp. 405-17.
- Groth, A. C. and Calos, M. P. (2004) 'Phage integrases: biology and applications', *J Mol Biol*, 335(3), pp. 667-78.
- Groth, A. C., Olivares, E. C., Thyagarajan, B. and Calos, M. P. (2000) 'A phage integrase directs efficient site-specific integration in human cells', *Proc Natl Acad Sci U S A*, 97(11), pp. 5995-6000.

- Halder, S., Banerjee, S. and Parrack, P. (2008) 'Direct CIII-HflB interaction is responsible for the inhibition of the HflB (FtsH)-mediated proteolysis of Escherichia coli sigma(32) by lambdaCIII', *FEBS J*, 275(19), pp. 4767-72.
- Hale, W. G., Saunders, V. A., Margham, J. P. and Hale, W. G. C. d. b. (2005) *Collins dictionary of biology*. Rev. and updated 3rd ed. edn. London: Collins.
- Hambly, E. and Suttle, C. A. (2005) 'The virosphere, diversity, and genetic exchange within phage communities', *Curr Opin Microbiol*, 8(4), pp. 444-50.
- Hauser, M. A. and Scocca, J. J. (1992) 'Site-specific integration of the Haemophilus influenzae bacteriophage HP1: location of the boundaries of the phage attachment site', *J Bacteriol*, 174(20), pp. 6674-7.
- Hendrix, R. W. (1983) *Lambda II. Cold Spring Harbor monograph series* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.
- Hendrix, R. W. and Duda, R. L. (1992) 'Bacteriophage lambda PaPa: not the mother of all lambda phages', *Science*, 258(5085), pp. 1145-8.
- Herman, C., Thévenet, D., D'Ari, R. and Bouloc, P. (1997) 'The HflB protease of Escherichia coli degrades its inhibitor lambda cIII', *J Bacteriol*, 179(2), pp. 358-63.
- Hirano, N., Muroi, T., Takahashi, H. and Haruki, M. (2011) 'Site-specific recombinases as tools for heterologous gene integration', *Appl Microbiol Biotechnol*, 92(2), pp. 227-39.
- Holt, G. S., Lodge, J. K., McCarthy, A. J., Graham, A. K., Young, G., Bridge, S. H., Brown, A. K., Veses-Garcia, M., Lanyon, C. V., Sails, A., Allison, H. E. and Smith, D. L. (2016) 'Shigatoxin encoding Bacteriophage ϕ 24B modulates bacterial metabolism to raise antimicrobial tolerance', *Scientific reports*, 7:40424, pp. 1-14.
- Hsu, P. L. and Landy, A. (1984) 'Resolution of synthetic att-site Holliday structures by the integrase protein of bacteriophage lambda', *Nature*, 311(5988), pp. 721-6.
- Huggett, J. F., Foy, C. A., Benes, V., Emslie, K., Garson, J. A., Haynes, R., Hellemans, J., Kubista, M., Mueller, R. D., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., Wittwer, C. T. and Bustin, S. A. (2013) 'The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments', *Clin Chem*, 59(6), pp. 892-902.
- Ikeda, H. and Tomizawa, J. (1968) 'Prophage P1, and extrachromosomal replication unit', *Cold Spring Harb Symp Quant Biol*, 33, pp. 791-8.
- Inoue, H., Nojima, H. and Okayama, H. (1990) 'High efficiency transformation of Escherichia coli with plasmids', *Gene*, 96(1), pp. 23-8.
- 'Isothermal Reaction (Gibson Assembly) Master Mix', (2017) *Cold Spring Harb Protoc*, 2017(3), pp. pdb.rec090019.

- Iwanowski, S. D. (Sept. 1892) 'Concerning the mosaic disease of the tobacco plant', *St. Petersburg Acad. Imp. Sci. Bul.*, 35, pp. 67-70.
- James, C. E., Stanley, K. N., Allison, H. E., Flint, H. J., Stewart, C. S., Sharp, R. J., Saunders, J. R. and McCarthy, A. J. (2001) 'Lytic and lysogenic infection of diverse *Escherichia coli* and *Shigella* strains with a verocytotoxigenic bacteriophage', *Appl Environ Microbiol*, 67(9), pp. 4335-7.
- Jayaram, M., Ma, C. H., Kachroo, A. H., Rowley, P. A., Guga, P., Fan, H. F. and Voziyanov, Y. (2015) 'An Overview of Tyrosine Site-specific Recombination: From an F1p Perspective', *Microbiol Spectr*, 3(4).
- Jonker, K. M., Tilburg, J. J., Hagele, G. H. and de Boer, E. (2008) 'Species identification in meat products using real-time PCR', *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*, 25(5), pp. 527-33.
- Kahmann, R., Rudt, F. and Mertens, G. (1984) 'Substrate and enzyme requirements for in vitro site-specific recombination in bacteriophage μ ', *Cold Spring Harb Symp Quant Biol*, 49, pp. 285-94.
- Kaiser, A. D. (1957) 'Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*', *Virology*, 3(1), pp. 42-61.
- Ke, S. H. and Madison, E. L. (1997) 'Rapid and efficient site-directed mutagenesis by single-tube 'megaprimer' PCR method', *Nucleic Acids Res*, 25(16), pp. 3371-2.
- Khalil, A. S. and Collins, J. J. (2010) 'Synthetic biology: applications come of age', *Nat Rev Genet*, 11(5), pp. 367-79.
- Kikuchi, Y. and Nash, H. A. (1978) 'The bacteriophage lambda int gene product. A filter assay for genetic recombination, purification of int, and specific binding to DNA', *J Biol Chem*, 253(20), pp. 7149-57.
- Kikuchi, Y. and Nash, H. A. (1979) 'Nicking-closing activity associated with bacteriophage lambda int gene product', *Proc Natl Acad Sci U S A*, 76(8), pp. 3760-4.
- Kilby, N. J., Snaith, M. R. and Murray, J. A. (1993) 'Site-specific recombinases: tools for genome engineering', *Trends Genet*, 9(12), pp. 413-21.
- Kitts, P. A. and Nash, H. A. (1987) 'Homology-dependent interactions in phage lambda site-specific recombination', *Nature*, 329(6137), pp. 346-8.
- Kobiler, O., Koby, S., Teff, D., Court, D. and Oppenheim, A. B. (2002) 'The phage lambda CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis', *Proc Natl Acad Sci U S A*, 99(23), pp. 14964-9.
- Kobiler, O., Rokney, A., Friedman, N., Court, D. L., Stavans, J. and Oppenheim, A. B. (2005) 'Quantitative kinetic analysis of the bacteriophage lambda genetic network', *Proc Natl Acad Sci U S A*, 102(12), pp. 4470-5.

- Kobiler, O., Rokney, A. and Oppenheim, A. B. (2007) 'Phage lambda CIII: a protease inhibitor regulating the lysis-lysogeny decision', *PLoS One*, 2(4), pp. e363.
- Kolot, M., Silberstein, N. and Yagil, E. (1999) 'Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022', *Mol Biol Rep*, 26(3), pp. 207-13.
- Koudelka, A. P., Hufnagel, L. A. and Koudelka, G. B. (2004) 'Purification and characterization of the repressor of the shiga toxin-encoding bacteriophage 933W: DNA binding, gene regulation, and autocleavage', *J Bacteriol*, 186(22), pp. 7659-69.
- Krasnow, M. A. and Cozzarelli, N. R. (1983) 'Site-specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented res sites', *Cell*, 32(4), pp. 1313-24.
- Kuhstoss, S. and Rao, R. N. (1991) 'Analysis of the integration function of the streptomycete bacteriophage phi C31', *J Mol Biol*, 222(4), pp. 897-908.
- Kwon, H. J., Tirumalai, R., Landy, A. and Ellenberger, T. (1997) 'Flexibility in DNA recombination: structure of the lambda integrase catalytic core', *Science*, 276(5309), pp. 126-31.
- Laemmli, U. K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227(5259), pp. 680-5.
- Landy, A. (2015) 'The λ Integrase Site-specific Recombination Pathway', *Microbiol Spectr*, 3(2), pp. MDNA3-0051-2014.
- Landy, A. and Ross, W. (1977) 'Viral integration and excision: structure of the lambda att sites', *Science*, 197(4309), pp. 1147-60.
- Law, D. (2000) 'Virulence factors of Escherichia coli O157 and other Shiga toxin-producing E. coli', *J Appl Microbiol*, 88(5), pp. 729-45.
- Lawrence, C. M., Menon, S., Eilers, B. J., Bothner, B., Khayat, R., Douglas, T. and Young, M. J. (2009) 'Structural and functional studies of archaeal viruses', *J Biol Chem*, 284(19), pp. 12599-603.
- Laxmikanthan, G., Xu, C., Brilot, A. F., Warren, D., Steele, L., Seah, N., Tong, W., Grigorieff, N., Landy, A. and Van Duyne, G. D. (2016) 'Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction', *Elife*, 5.
- Lee, C. Y., Buranen, S. L. and Ye, Z. H. (1991a) 'Construction of single-copy integration vectors for Staphylococcus aureus', *Gene*, 103(1), pp. 101-5.
- Lee, M. H. and Hatfull, G. F. (1993) 'Mycobacteriophage L5 integrase-mediated site-specific integration in vitro', *J Bacteriol*, 175(21), pp. 6836-41.
- Lee, M. H., Pascopella, L., Jacobs, W. R. and Hatfull, G. F. (1991b) 'Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guérin', *Proc Natl Acad Sci U S A*, 88(8), pp. 3111-5.

- Lee, T. S., Krupa, R. A., Zhang, F., Hajimorad, M., Holtz, W. J., Prasad, N., Lee, S. K. and Keasling, J. D. (2011) 'BglBrick vectors and datasheets: A synthetic biology platform for gene expression', *J Biol Eng*, 5, pp. 12.
- Leong, J. M., Nunes-Düby, S., Lesser, C. F., Youderian, P., Susskind, M. M. and Landy, A. (1985) 'The phi 80 and P22 attachment sites. Primary structure and interaction with Escherichia coli integration host factor', *J Biol Chem*, 260(7), pp. 4468-77.
- Le, S., Chen, H., Cong, P., Lin, J., Dröge, P. and Yan, J. (2013) 'Mechanosensing of DNA bending in a single specific protein-DNA complex', *Sci Rep*, 3, pp. 3508.
- Lewis, C. T., Short, C., Freund, W. and Andrews, E. A. (1969) *A Latin dictionary founded on Andrews' edition of Freund's Latin dictionary*. Oxford: Clarendon Press.
- Lewis, D. E., Gussin, G. N. and Adhya, S. (2016) 'New Insights into the Phage Genetic Switch: Effects of Bacteriophage Lambda Operator Mutations on DNA Looping and Regulation of PR, PL, and PRM', *J Mol Biol*, 428(22), pp. 4438-4456.
- Lewis, J. A. and Hatfull, G. F. (2000) 'Identification and characterization of mycobacteriophage L5 excisionase', *Mol Microbiol*, 35(2), pp. 350-60.
- Lewis, J. A. and Hatfull, G. F. (2001) 'Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins', *Nucleic Acids Res*, 29(11), pp. 2205-16.
- Lewis, J. A. and Hatfull, G. F. (2003) 'Control of directionality in L5 integrase-mediated site-specific recombination', *J Mol Biol*, 326(3), pp. 805-21.
- Liang, X., Peng, L., Baek, L. and Katzen, F. (2013) 'Single step BP/LR combined Gateway reactions', *BioTechniques*, 55, pp. 265-268.
- Lingwood, C. A. (1999) 'Glycolipid receptors for verotoxin and Helicobacter pylori: role in pathology', *Biochim Biophys Acta*, 1455(2-3), pp. 375-86.
- Lisio, A. L. and Weissbach, A. (1965) 'Repression of lambda-Associated Enzyme Synthesis After lambda(vir) Superinfection of Lysogenic Hosts', *J Bacteriol*, 90(3), pp. 661-6.
- Liu, J. K., Chen, W. H., Ren, S. X., Zhao, G. P. and Wang, J. (2014) 'iBrick: a new standard for iterative assembly of biological parts with homing endonucleases', *PLoS One*, 9(10), pp. e110852.
- Lorbach, E., Christ, N., Schwikardi, M. and Droge, P. (2000) 'Site-specific recombination in human cells catalyzed by phage lambda integrase mutants', *J Mol Biol*, 296(5), pp. 1175-81.
- Madsen, P. L., Johansen, A. H., Hammer, K. and Brøndsted, L. (1999) 'The genetic switch regulating activity of early promoters of the temperate lactococcal bacteriophage TP901-1', *J Bacteriol*, 181(24), pp. 7430-8.

- Mattis, A. N., Gumpert, R. I. and Gardner, J. F. (2008) 'Purification and characterization of bacteriophage P22 Xis protein', *J Bacteriol*, 190(17), pp. 5781-96.
- McCulloch, R., Coggins, L. W., Colloms, S. D. and Sherratt, D. J. (1994) 'Xer-mediated site-specific recombination at *cer* generates Holliday junctions in vivo', *EMBO J*, 13(8), pp. 1844-55.
- McGrath, S., Fitzgerald, G. F. and van Sinderen, D. (2002) 'Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages', *Mol Microbiol*, 43(2), pp. 509-20.
- Meselson, M. and Weigle, J. J. (1961) 'Chromosome breakage accompanying genetic recombination in bacteriophage', *Proc Natl Acad Sci U S A*, 47, pp. 857-68.
- Miyazaki, R. and van der Meer, J. R. (2013) 'A new large-DNA-fragment delivery system based on integrase activity from an integrative and conjugative element', *Appl Environ Microbiol*, 79(14), pp. 4440-7.
- Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., Foeller, C., Hsu, P. L., Ross, W. and Landy, A. (1981) 'Structure and function of the phage lambda att site: size, int-binding sites, and location of the crossover point', *Cold Spring Harb Symp Quant Biol*, 45 Pt 1, pp. 429-37.
- Mizuuchi, M. and Mizuuchi, K. (1980) 'Integrative recombination of bacteriophage lambda: extent of the DNA sequence involved in attachment site function', *Proc Natl Acad Sci U S A*, 77(6), pp. 3220-4.
- Morita, K., Yamamoto, T., Fusada, N., Komatsu, M., Ikeda, H., Hirano, N. and Takahashi, H. (2009) 'The site-specific recombination system of actinophage TG1', *FEMS Microbiol Lett*, 297(2), pp. 234-40.
- Mustard, J. A. and Little, J. W. (2000) 'Analysis of Escherichia coli RecA interactions with LexA, lambda CI, and UmuD by site-directed mutagenesis of *recA*', *J Bacteriol*, 182(6), pp. 1659-70.
- Nagaraja, R. and Weisberg, R. A. (1990) 'Specificity determinants in the attachment sites of bacteriophages HK022 and lambda', *J Bacteriol*, 172(11), pp. 6540-50.
- Nash, H. A. (1990) 'Bending and supercoiling of DNA at the attachment site of bacteriophage lambda', *Trends Biochem Sci*, 15(6), pp. 222-7.
- Nash, H. A. and Pollock, T. J. (1983) 'Site-specific recombination of bacteriophage lambda. The change in topological linking number associated with exchange of DNA strands', *J Mol Biol*, 170(1), pp. 19-38.
- Nataro, J. P. and Kaper, J. B. (1998) 'Diarrheagenic Escherichia coli', *Clin Microbiol Rev*, 11(1), pp. 142-201.
- Novick, R. P., Clowes, R. C., Cohen, S. N., Curtiss, R., Datta, N. and Falkow, S. (1976) 'Uniform nomenclature for bacterial plasmids: a proposal', *Bacteriol Rev*, 40(1), pp. 168-89.

- Numrych, T. E., Gumport, R. I. and Gardner, J. F. (1992) 'Characterization of the bacteriophage lambda excisionase (Xis) protein: the C-terminus is required for Xis-integrase cooperativity but not for DNA binding', *EMBO J*, 11(10), pp. 3797-806.
- Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. and Landy, A. (1998) 'Similarities and differences among 105 members of the Int family of site-specific recombinases', *Nucleic Acids Res*, 26(2), pp. 391-406.
- Nunes-Düby, S. E., Matsumoto, L. and Landy, A. (1989) 'Half-att site substrates reveal the homology independence and minimal protein requirements for productive synapsis in lambda excisive recombination', *Cell*, 59(1), pp. 197-206.
- O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W. and Formal, S. B. (1984) 'Shiga-like toxin-converting phages from Escherichia coli strains that cause hemorrhagic colitis or infantile diarrhea', *Science*, 226(4675), pp. 694-6.
- Olivares, E. C., Hollis, R. P. and Calos, M. P. (2001) 'Phage R4 integrase mediates site-specific integration in human cells', *Gene*, 278(1-2), pp. 167-76.
- Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L. and Adhya, S. (2005) 'Switches in bacteriophage lambda development', *Annu Rev Genet*, 39, pp. 409-29.
- Park, J., Throop, A. L. and LaBaer, J. (2015) 'Site-specific recombinational cloning using gateway and in-fusion cloning schemes', *Curr Protoc Mol Biol*, 110, pp. 3.20.1-23.
- Paton, A. W. and Paton, J. C. (1999) 'Direct detection of Shiga toxigenic Escherichia coli strains belonging to serogroups O111, O157, and O113 by multiplex PCR', *J Clin Microbiol*, 37(10), pp. 3362-5.
- Paton, A. W. and Paton, J. C. (1998) 'Detection and characterization of Shiga toxigenic Escherichia coli by using multiplex PCR assays for stx1, stx2, eaeA, enterohemorrhagic E. coli hlyA, rfb0111, and rfb0157', *J Clin Microbiol*, 36(2), pp. 598-602.
- Patsey, R. L. and Bruist, M. F. (1995) 'Characterization of the interaction between the lambda intasome and attB', *J Mol Biol*, 252(1), pp. 47-58.
- Pedersen, M., Ligowska, M. and Hammer, K. (2010) 'Characterization of the CI repressor protein encoded by the temperate lactococcal phage TP901-1', *J Bacteriol*, 192(8), pp. 2102-10.
- Peixoto, A., Monteiro, M., Rocha, B. and Veiga-Fernandes, H. (2004) 'Quantification of multiple gene expression in individual cells', *Genome Res*, 14(10A), pp. 1938-47.
- Prescott, L. M., Harley, J. P. and Klein, D. A. (eds.) (2002) *Microbiology*. 5th edn. Avenue of the Americas, New York, NY, USA: McGraw-Hill Companies, Inc.

- Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Dürr, F., Emenegger, J., Kelterborn, S., Sauer, C., Fritz, G., Gebhard, S. and Mascher, T. (2013) 'The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*', *J Biol Eng*, 7(1), pp. 29.
- Rajaure, M., Berry, J., Kongari, R., Cahill, J. and Young, R. (2015) 'Membrane fusion during phage lysis', *Proc Natl Acad Sci U S A*, 112(17), pp. 5497-502.
- Rajeev, L., Malanowska, K. and Gardner, J. F. (2009) 'Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions', *Microbiol Mol Biol Rev*, 73(2), pp. 300-9.
- Rankin, S. and Platt, D. J. (1995) 'Phage conversion in *Salmonella enterica* serotype Enteritidis: implications for epidemiology', *Epidemiol Infect*, 114(2), pp. 227-36.
- Raya, R. R. and H'bert, E. M. (2009) 'Isolation of Phage via Induction of Lysogens', *Methods Mol Biol*, 501, pp. 23-32.
- Reed, R. R., Shibuya, G. I. and Steitz, J. A. (1982) 'Nucleotide sequence of gamma delta resolvase gene and demonstration that its gene product acts as a repressor of transcription', *Nature*, 300(5890), pp. 381-3.
- Richet, E., Abcarian, P. and Nash, H. A. (1988) 'Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex', *Cell*, 52(1), pp. 9-17.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Séraphin, B. (1999) 'A generic protein purification method for protein complex characterization and proteome exploration', *Nat Biotechnol*, 17(10), pp. 1030-2.
- Riley, L. M., Veses-Garcia, M., Hillman, J. D., Handfield, M., McCarthy, A. J. and Allison, H. E. (2012) 'Identification of genes expressed in cultures of *E. coli* lysogens carrying the Shiga toxin-encoding prophage Φ 24B', *BMC Microbiol*, 12, pp. 42.
- Rist, M. and Kertesz, M. A. (1998) 'Construction of improved plasmid vectors for promoter characterization in *Pseudomonas aeruginosa* and other gram-negative bacteria', *FEMS Microbiol Lett*, 169(1), pp. 179-83.
- Ritchie, D. A. and Joicey, D. H. (1978) 'Formation of concatemeric DNA as an intermediate in the replication of bacteriophage T1 DNA molecules', *J Gen Virol*, 41(3), pp. 609-22.
- Roberts, J. W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D. C., Sun, H. and Roberts, C. W. (1998) 'Antitermination by bacteriophage lambda Q protein', *Cold Spring Harb Symp Quant Biol*, 63, pp. 319-25.
- Robles-Oteiza, C., Taylor, S., Yates, T., Cicchini, M., Lauderback, B., Cashman, C. R., Burds, A. A., Winslow, M. M., Jacks, T. and Feldser, D. M. (2015)

- 'Recombinase-based conditional and reversible gene regulation via XTR alleles', *Nat Commun*, 6, pp. 8783.
- Rodriguez, E., Hu, Z., Ou, S., Volchegursky, Y., Hutchinson, C. R. and McDaniel, R. (2003) 'Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains', *J Ind Microbiol Biotechnol*, 30(8), pp. 480-8.
- Rooks, D. J., Yan, Y., McDonald, J. E., Woodward, M. J., McCarthy, A. J. and Allison, H. E. (2010) 'Development and validation of a qPCR-based method for quantifying Shiga toxin-encoding and other lambdoid bacteriophages', *Environ Microbiol*, 12(5), pp. 1194-204.
- Ross, W., Landy, A., Kikuchi, Y. and Nash, H. (1979) 'Interaction of int protein with specific sites on lambda att DNA', *Cell*, 18(2), pp. 297-307.
- Rutkai, E., Dorgai, L., Sirot, R., Yagil, E. and Weisberg, R. A. (2003) 'Analysis of insertion into secondary attachment sites by phage lambda and by int mutants with altered recombination specificity', *J Mol Biol*, 329(5), pp. 983-96.
- Røkke, G., Korvald, E., Pahr, J., Oyås, O. and Lale, R. (2014) 'BioBrick assembly standards and techniques and associated software tools', *Methods Mol Biol*, 1116, pp. 1-24.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual SECOND EDITION*. 2th edition edn. (1 vols). 10 Skyline drive, Plainview, New York, USA.: Cold Spring Harbor Laboratory Press.
- Sanchez, J. and Holmgren, J. (2011) 'Cholera toxin - a foe & a friend', *Indian J Med Res*, 133, pp. 153-63.
- Sandvig, K. (2001) 'Shiga toxins', *Toxicon*, 39(11), pp. 1629-35.
- Sanyal, S. J., Yang, T. C. and Catalano, C. E. (2014) 'Integration host factor assembly at the cohesive end site of the bacteriophage lambda genome: implications for viral DNA packaging and bacterial gene regulation', *Biochemistry*, 53(48), pp. 7459-70.
- Sarkar, T., Petrov, A. S., Vitko, J. R., Santai, C. T., Harvey, S. C., Mukerji, I. and Hud, N. V. (2009) 'Integration host factor (IHF) dictates the structure of polyamine-DNA condensates: implications for the role of IHF in the compaction of bacterial chromatin', *Biochemistry*, 48(4), pp. 667-75.
- Sauer, B. (1998) 'Inducible gene targeting in mice using the Cre/lox system', *Methods*, 14(4), pp. 381-92.
- Saunders, J., Allison, H., James, C., McCarthy, A. and Sharp, R. (2001) 'Phage-mediated transfer of virulence genes', *J Chem Tech Biotech*, 76(7), pp. 662-666.
- Schindler, D. and Echols, H. (1981) 'Retroregulation of the int gene of bacteriophage lambda: control of translation completion', *Proc Natl Acad Sci USA*, 78(7), pp. 4475-9.

- Schmidt, S. D., Nixon, R. A. and Mathews, P. M. (2012) 'Tissue processing prior to analysis of Alzheimer's disease associated proteins and metabolites, including A β ', *Methods Mol Biol*, 849, pp. 493-506.
- Serra-Moreno, R., Jofre, J. and Muniesa, M. (2008) 'The CI repressors of Shiga toxin-converting prophages are involved in coinfection of Escherichia coli strains, which causes a down regulation in the production of Shiga toxin 2', *J Bacteriol*, 190(13), pp. 4722-35.
- Shaw, N., Lehner, B., Fuhrmann, M., Kulla, H., Brass, J., Birch, O., Tinschert, A., Venetz, D., Venetz, V., Sanchez, J. C., Tonella, L. and Hochstrasser, D. (1999) 'Biotin production under limiting growth conditions by Agrobacterium/Rhizobium HK4 transformed with a modified Escherichia coli bio operon', *J Ind Microbiol Biotechnol*, 22(6), pp. 590-599.
- Shetty, R. P., Endy, D. and Knight, T. F. (2008) 'Engineering BioBrick vectors from BioBrick parts', *J Biol Eng*, 2, pp. 5.
- Shevchuk, N. A., Bryksin, A. V., Nusinovich, Y. A., Cabello, F. C., Sutherland, M. and Ladisch, S. (2004) 'Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously', *Nucleic Acids Res*, 32(2), pp. e19.
- Shotland, Y., Koby, S., Teff, D., Mansur, N., Oren, D. A., Tatematsu, K., Tomoyasu, T., Kessel, M., Bukau, B., Ogura, T. and Oppenheim, A. B. (1997) 'Proteolysis of the phage lambda CII regulatory protein by FtsH (HflB) of Escherichia coli', *Mol Microbiol*, 24(6), pp. 1303-10.
- Shpakovski, G. V. and Berlin, Y. A. (1984) 'Site-specificity of abnormal excision: the mechanism of formation of a specialized transducing bacteriophage lambda plac5', *Nucleic Acids Res*, 12(17), pp. 6779-95.
- Siegel, L. M. and Monty, K. J. (1966) 'Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases', *Biochim Biophys Acta*, 112(2), pp. 346-62.
- Signer, E. R. and Weil, J. (1968) 'Site-specific recombination in bacteriophage lambda', *Cold Spring Harb Symp Quant Biol*, 33, pp. 715-9.
- Sleight, S. C., Bartley, B. A., Lieviant, J. A. and Sauro, H. M. (2010) 'In-Fusion BioBrick assembly and re-engineering', *Nucleic Acids Res*, 38(8), pp. 2624-36.
- Smith, H. O. and Wilcox, K. W. (1970) 'A restriction enzyme from Hemophilus influenzae. I. Purification and general properties', *J Mol Biol*, 51(2), pp. 379-91.
- Smith, M. C. and Thorpe, H. M. (2002) 'Diversity in the serine recombinases', *Mol Microbiol*, 44(2), pp. 299-307.

- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. and Klenk, D. C. (1985) 'Measurement of protein using bicinchoninic acid', *Anal Biochem*, 150(1), pp. 76-85.
- Smith-Mungo, L., Chan, I. T. and Landy, A. (1994) 'Structure of the P22 att site. Conservation and divergence in the lambda motif of recombinogenic complexes', *J Biol Chem*, 269(32), pp. 20798-805.
- St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B. and Shearwin, K. E. (2013) 'One-step cloning and chromosomal integration of DNA', *ACS Synth Biol*, 2(9), pp. 537-41.
- Stark, W. M., Boocock, M. R. and Sherratt, D. J. (1992) 'Catalysis by site-specific recombinases', *Trends Genet*, 8(12), pp. 432-9.
- Stodolsky, M., Rae, M. E. and Mullenbach, E. (1972) 'The addition of lac+ chromosome fragments to the E. coli proA-proB-lac deletion 13 chromosome', *Genetics*, 70(4), pp. 495-510.
- Stoll, S. M., Ginsburg, D. S. and Calos, M. P. (2002) 'Phage TP901-1 site-specific integrase functions in human cells', *J Bacteriol*, 184(13), pp. 3657-63.
- Stover, C. K., de la Cruz, V. F., Fuerst, T. R., Burlein, J. E., Benson, L. A., Bennett, L. T., Bansal, G. P., Young, J. F., Lee, M. H. and Hatfull, G. F. (1991) 'New use of BCG for recombinant vaccines', *Nature*, 351(6326), pp. 456-60.
- Sugimura, S. and Crothers, D. M. (2006) 'Stepwise binding and bending of DNA by Escherichia coli integration host factor', *Proc Natl Acad Sci U S A*, 103(49), pp. 18510-4.
- Summer, E. J., Berry, J., Tran, T. A., Niu, L., Struck, D. K. and Young, R. (2007) 'Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts', *J Mol Biol*, 373(5), pp. 1098-112.
- Sylwan, L., Frumerie, C. and Haggård-Ljungquist, E. (2010) 'Identification of bases required for P2 integrase core binding and recombination', *Virology*, 404(2), pp. 240-5.
- Thomason, L. C., Calendar, R. and Ow, D. W. (2001) 'Gene insertion and replacement in Schizosaccharomyces pombe mediated by the Streptomyces bacteriophage phiC31 site-specific recombination system', *Mol Genet Genomics*, 265(6), pp. 1031-8.
- Thompson, J. F., de Vargas, L. M., Skinner, S. E. and Landy, A. (1987) 'Protein-protein interactions in a higher-order structure direct lambda site-specific recombination', *J Mol Biol*, 195(3), pp. 481-93.
- Thorpe, H. M. and Smith, M. C. (1998) 'In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family', *Proc Natl Acad Sci U S A*, 95(10), pp. 5505-10.

- Thorpe, H. M., Wilson, S. E. and Smith, M. C. (2000) 'Control of directionality in the site-specific recombination system of the *Streptomyces* phage phiC31', *Mol Microbiol*, 38(2), pp. 232-41.
- Tirumalai, R. S., Kwon, H. J., Cardente, E. H., Ellenberger, T. and Landy, A. (1998) 'Recognition of core-type DNA sites by lambda integrase', *J Mol Biol*, 279(3), pp. 513-27.
- Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. and Calos, M. P. (2001) 'Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase', *Mol Cell Biol*, 21(12), pp. 3926-34.
- Van Duyne, G. D. (2005) 'Lambda integrase: armed for recombination', *Curr Biol*, 15(17), pp. R658-60.
- Van den Ent, F. and Löwe, J. (2006) 'RF cloning: a restriction-free method for inserting target genes into plasmids', *J Biochem Biophys Methods*, 67(1), pp. 67-74.
- Veses-Garcia, M., Liu, X., Rigden, D. J., Kenny, J. G., McCarthy, A. J. and Allison, H. E. (2015) 'Transcriptomic analysis of Shiga-toxigenic bacteriophage carriage reveals a profound regulatory effect on acid resistance in *Escherichia coli*', *Appl Environ Microbiol*, 81(23), pp. 8118-25.
- Voigt, C. (2011) *Synthetic biology*. Amsterdam: Elsevier.
- Waldman, A. S., Fitzmaurice, W. P. and Scocca, J. J. (1986) 'Integration of the bacteriophage HP1c1 genome into the *Haemophilus influenzae* Rd chromosome in the lysogenic state', *J Bacteriol*, 165(1), pp. 297-300.
- Waldor, M. K. and Mekalanos, J. J. (1996) 'Lysogenic conversion by a filamentous phage encoding cholera toxin', *Science*, 272(5270), pp. 1910-4.
- Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S. and Vidal, M. (2000) 'GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes', *Methods Enzymol*, 328, pp. 575-592.
- Wang, Z., Xiong, G. and Lutz, F. (1995) 'Site-specific integration of the phage phi CTX genome into the *Pseudomonas aeruginosa* chromosome: characterization of the functional integrase gene located close to and upstream of attP', *Mol Gen Genet*, 246(1), pp. 72-9.
- Watanabe, Y., Ozasa, K., Mermin, J. H., Griffin, P. M., Masuda, K., Imashuku, S. and Sawada, T. (1999) 'Factory outbreak of *Escherichia coli* O157:H7 infection in Japan', *Emerg Infect Dis*, 5(3), pp. 424-8.
- Weickert, M. J. and Adhya, S. (1993) 'The galactose regulon of *Escherichia coli*', *Mol Microbiol*, 10(2), pp. 245-51.
- Weinbauer, M. G. (2004) 'Ecology of prokaryotic viruses', *FEMS Microbiol Rev*, 28(2), pp. 127-81.

- Weisberg, R. A., Enquist, L. W., Foeller, C. and Landy, A. (1983) 'Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda', *J Mol Biol*, 170(2), pp. 319-42.
- Wiechelman, K. J., Braun, R. D. and Fitzpatrick, J. D. (1988) 'Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation', *Anal Biochem*, 175(1), pp. 231-7.
- Wilhelm, S. W. and Suttle, C. A. (1999) 'Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs', *BioScience*, 49(10), pp. 781-788.
- Wilson, J. H. and Hunt, T. (2002) *Molecular biology of the cell, 4th edition : a problems approach*. New York ; London: Garland Science.
- Wingfield, P. (2001) 'Protein precipitation using ammonium sulfate', *Curr Protoc Protein Sci*, Appendix 3, pp. Appendix 3F.
- Xu, X., McAteer, S. P., Tree, J. J., Shaw, D. J., Wolfson, E. B., Beatson, S. A., Roe, A. J., Allison, L. J., Chase-Topping, M. E., Mahajan, A., Tozzoli, R., Woolhouse, M. E., Morabito, S. and Gally, D. L. (2012) 'Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic Escherichia coli', *PLoS Pathog*, 8(5), pp. e1002672.
- Yang, H. Y., Kim, Y. W. and Chang, H. I. (2002) 'Construction of an integration-proficient vector based on the site-specific recombination mechanism of enterococcal temperate phage phiFC1', *J Bacteriol*, 184(7), pp. 1859-64.
- Yang, S. W. and Nash, H. A. (1995) 'Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target', *EMBO J*, 14(24), pp. 6292-300.
- Yang, X. J., Hart, C. M., Grayhack, E. J. and Roberts, J. W. (1987) 'Transcription antitermination by phage lambda gene Q protein requires a DNA segment spanning the RNA start site', *Genes Dev*, 1(3), pp. 217-26.
- Yin, S., Bushman, W. and Landy, A. (1985) 'Interaction of the lambda site-specific recombination protein Xis with attachment site DNA', *Proc Natl Acad Sci U S A*, 82(4), pp. 1040-4.
- Young, R., Way, J., Way, S., Yin, J. and Syvanen, M. (1979) 'Transposition mutagenesis of bacteriophage lambda: a new gene affecting cell lysis', *J Mol Biol*, 132(3), pp. 307-22.
- Yu, A. and Haggård-Ljungquist, E. (1993) 'The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination', *J Bacteriol*, 175(24), pp. 7848-55.
- Yue, W. F., Du, M. and Zhu, M. J. (2012) 'High temperature in combination with UV irradiation enhances horizontal transfer of stx2 gene from E. coli O157:H7 to non-pathogenic E. coli', *PLoS One*, 7(2), pp. e31308.
- Zhang, B., Zhang, L., Dai, R., Yu, M., Zhao, G. and Ding, X. (2013) 'An efficient procedure for marker-free mutagenesis of *S. coelicolor* by site-specific

recombination for secondary metabolite overproduction', *PLoS One*, 8(2), pp. e55906.

Zhang, G., Fenyő, D. and Neubert, T. A. (2008a) 'Use of DNA ladders for reproducible protein fractionation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) for quantitative proteomics', *J Proteome Res*, 7(2), pp. 678-86.

Zhang, L., Ou, X., Zhao, G. and Ding, X. (2008b) 'Highly efficient in vitro site-specific recombination system based on streptomyces phage phiBT1 integrase', *J Bacteriol*, 190(19), pp. 6392-7.